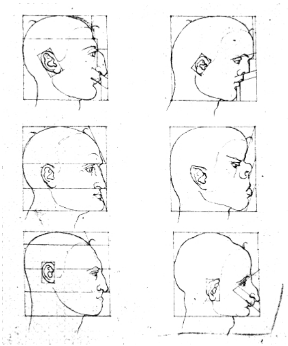
Statistical inference of facial beauty

Miguel Ibáñez Berganza

Lipari School, July 15^{th}, 2019

SAPIENZA
UNIVERSITÀ DI ROMA



[Durer, Four books on human proportion 1534]

Accuracy/complexity trade-off (model selection) in statistical inference

with applications in cognitive science, neuroscience and urbanism

Accuracy/complexity trade-off (model selection) in statistical inference

with applications in cognitive science, neuroscience and urbanism

- Overfitting regularisation in Maximum Entropy inference [A. Gabrielli, T. Gili, F. Santucci]

Accuracy/complexity trade-off (model selection) in statistical inference

with applications in cognitive science, neuroscience and urbanism

- Overfitting regularisation in Maximum Entropy inference [A. Gabrielli, T. Gili, F. Santucci]
- Statistical inference and the relation among **brain function and structure**
  [A. Gabrielli, T. Gili, F. Santucci]
  [MaxEnt & RBM inference of human brain fMRI time series]

Accuracy/complexity trade-off (model selection) in statistical inference

with applications in cognitive science, neuroscience and urbanism

- Overfitting regularisation in Maximum Entropy inference [A. Gabrielli, T. Gili, F. Santucci]
- Statistical inference and the relation among **brain function and structure**
  [A. Gabrielli, T. Gili, F. Santucci]
  [MaxEnt & RBM inference of human brain fMRI time series]
- Forecasting the temporal evolution of cities
  [Monechi+ 2019?]
  [Prediction of the evolution of urban socio-economical indicators from past-year values]

Accuracy/complexity trade-off (model selection) in statistical inference

with applications in cognitive science, neuroscience and urbanism

- Overfitting regularisation in Maximum Entropy inference [A. Gabrielli, T. Gili, F. Santucci]
- Statistical inference and the relation among **brain function and structure** [A. Gabrielli, T. Gili, F. Santucci]
  [MaxEnt & RBM inference of human brain fMRI time series]
- Forecasting the temporal evolution of cities
  [Monechi+ 2019?]
  [Prediction of the evolution of urban socio-economical indicators from past-year values]
- Statistical inference analysis of the human facial perception process
  [I-B+ 2019]; [I-B+ 2019?]
  [Assessment of the subjectivity of the facial preference and of the relevant facial features involved in the cognitive process]

Unsupervised inference of facial attractiveness

[I.-B., Amico, Loreto, Sci. Rep. **9** 8364, 2019]

[I.-B., Monechi, Lancia, Loreto, preprint 2019]

Quantitative research in facial attractiveness

- Evolutionary biology: validity of the "health certificate hypothesis"

Quantitative research in facial attractiveness

- Evolutionary biology: validity of the “health certificate hypothesis”
    - some extent of observed universality
    - assessment of covariation between attractiveness and facial traits believed to signal health/fertility: **symmetry**, **averageness** and **sexually dimorphic features**
    - ...whose impact results rather controversial, and limited

Quantitative research in facial attractiveness

- Evolutionary biology: validity of the "health certificate hypothesis"
  - some extent of observed universality
  - assessment of covariation between attractiveness and facial traits believed to signal health/fertility:
    symmetry, **averageness** and **sexually dimorphic features**
  - ...whose impact results rather controversial, and limited
  - studies are mainly based on *average ratings* assigned by volunteers to *natural facial images*

Quantitative research in facial attractiveness

- Machine Learning: automatic rating of facial images

Quantitative research in facial attractiveness

- Machine Learning: automatic rating of facial images
  - supervised inference of the average rating

    database: i-th facial image $\rightarrow \mathbf{v}_i$, ratings: $r_i^{(s)}$

    inference: $R(\mathbf{v}_i|\boldsymbol{\theta}) \sim \langle r_i^{(s)} \rangle_s$

Quantitative research in facial attractiveness

- Machine Learning: automatic rating of facial images
  - supervised inference of the average rating

    database: i-th facial image $\rightarrow \mathbf{v}_i$, ratings: $r_i^{(s)}$

    inference: $R(\mathbf{v}_i|\boldsymbol{\theta}) \sim \langle r_i^{(s)} \rangle_s$
  - hyerarchical ANN's: the raw image is passed to the algorithm (suffering from the *black box problem*)

Quantitative research in facial attractiveness

Technique	Database				Score levels	Panel size	Facial features used	Classification technique / beauty predictor	Validation method	Accuracy
	Size	Gender	Expr.	Color						
							smoothness, hair color; feature variables were decorrelated with PCA (2) selected eigenfaces	with KNN and SVM (b) Beauty predictor on 7 classes with KNN, Linear Regression (LR), SVM		predictor scores 0.65 as correlation
Chen et al. [28]	100	F	N	N	10	18	Gabor features, reduced by PCA, +Kagian 2008 features	Support Vector Regression and KNN, Linear Regression + Feature selection	Leave-one-out cross validation	0.93 Pearson correlation (0.89 using Gabor features only)
Turkmen et al. [157]	160	F	N	N	2	50	Eigenfaces (computed with PCA and KPCA)	SVM	Training 90, testing 70	82.5% correct classification with PCA, 88.75% with KPCA
Bronstad et al. [21]	74	F/M	N	N	7	102	Partial least square, Geometric feature reduced by PCA	Perceptron	Leave-one-out cross validation	Pearson correlations PLS: 0.7 for female faces, 0.68 for male faces Geometric features + PCA: 0.78 for male faces and 0.61 for female faces
Said and Todorov [136]	4200 (synthetic)	F/M	N	Y	9	40	25 shape and 25 reflectivity features, both computed with PCA	Non-linear regression	4000 training, 200 testing	$R^2 = 0.79$ for female faces and $R^2 = 0.84$ for male faces
Internet DB										
Sutic et al. [150]	Set 1: 136 Set 2: 200	F/M	Y	Y	10	>50	Set 1: 25 geometric ratios Set 2: an unspecified number of eigenfaces	Classification with KNN, Neural Networks, AdaBoost in two experiments: (a) 2 class classification (class separation given by the median of all scores) (b) 4 class classification (boundaries are quartiles of all values)	Set 1: 70 for training, 30 for validation and 36 as test set. Set 2: 100 for training, 100 for testing	(a) 67% correct classification in the best case (KNN with eigenfaces) (b) 33% of correct classification with KNN (the feature set was not specified)
Gray et al. [62]	2056	F	Y	Y	N/A	30	Eigenfaces, multiscale single layer local filters, single and two-layer local filters	Regression model Note: score levels are continuous, recomputed from pairwise ratings	1028 for training, 1028 for testing	Top correlation with recomputed ratings: 0.458 for multiscale model
Dantcheva and Dugelay [33]	325	F	Y	Y	10	>50	Landmark locations, geometric ratios, geometric features, expressions, non-permanent traits image attributes	Multiple regression	260 for training, 65 for testing	0.77 Pearson correlation
Whithehill and Movellan [169]	2000	F/M	Y	N	4	8	Gabor features, eigenfaces, geometric features, Edge Orientation Histograms (EOH)	ε-SVM regression	5-fold cross validation on the dataset chosen by each rater	0.28 correlation with personal preferences using Gabor features, 0.26 with PCA, 0.24 using EOH, while geometric features scored only a 0.08 correlation
Altwaijry and Belongie [6]	200	F	N	Y	N/A	60	Geometric features, HOG, GIST, $L*a*b$ color histograms, eigenfaces, SIFT, Dense-SIFT reduced with PCA	Rank learning based on a modified SVM approach	160 training, 40 testing	63% accuracy, obtained combining all features except eigenfaces

[Laurentini+ 2014]

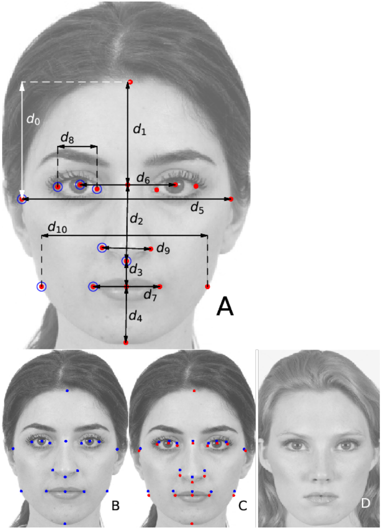
Unanswered questions

- Which are the suitable facial variables for an efficient description of the phenomenon?
  (texture and geometric degrees of freedom are probably **coupled**)
- What is the extent/origin of the inter-subject diversity?

A novel experimental method

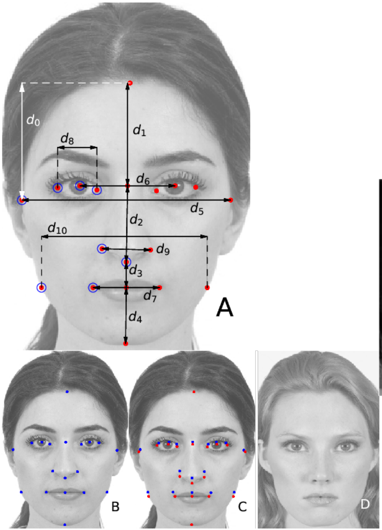
- Reduction of the face space dimension (only geometric quantities vary)
- No rating (the subject explores the face-space instead)

Facial preference I: subjectivity of attractiveness perception



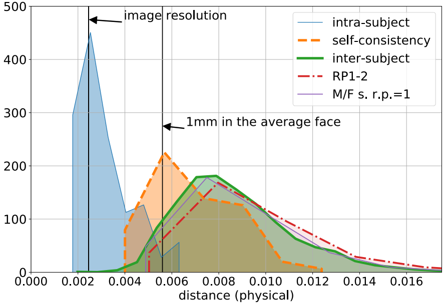
$$\mathcal{S} = \{\mathbf{x}^{(s)}\}_{s=1}^{S}$$

Facial preference I: subjectivity of attractiveness perception

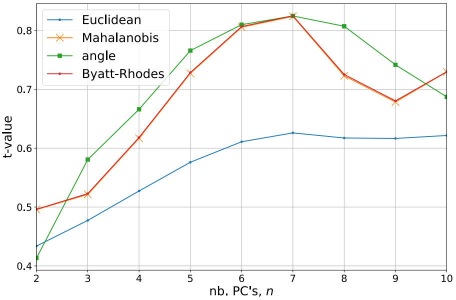


$$\mathcal{S} = \{\mathbf{x}^{(s)}\}_{s=1}^{S}$$

Facial preference I: subjectivity of attractiveness perception

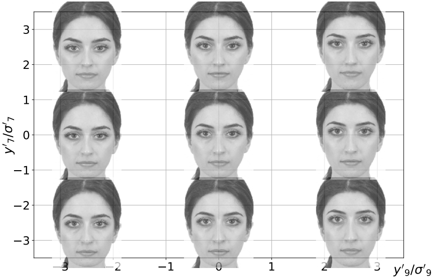


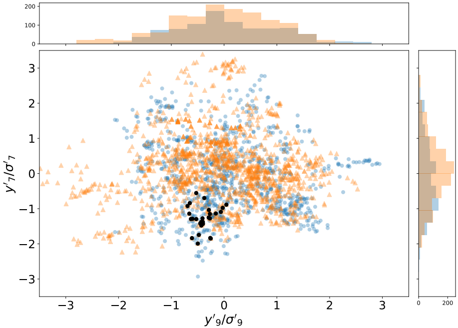
Facial preference I: subjectivity of attractiveness perception



$$\mathrm{t} = (\mu_{\mathrm{is}} - \mu_{\mathrm{sc}})/(\sigma_{\mathrm{is}}^2 + \sigma_{\mathrm{sc}}^2)^{1/2}$$

Facial preference I: subjectivity of attractiveness perception

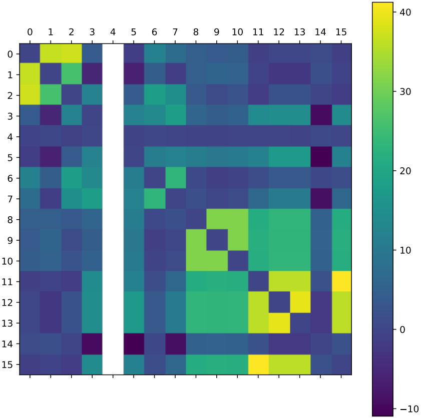




Facial preference II: inference of the database of sculpted vectors

$\vec{\boldsymbol{\Delta}} = (\vec{\Delta}_i)_i$ fluctuations of the i-th landmark Cartesian coordinates wrt. their averages

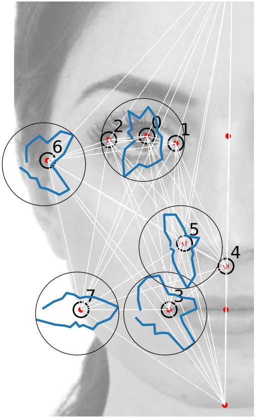
The fluctuations are highly correlated



$$C_{\alpha\beta} = \langle \Delta_\alpha \Delta_\beta \rangle,\ \alpha = (i, \mathrm{c} = (\mathrm{x}, \mathrm{y}))$$

Facial preference II: inference of the database of sculpted vectors

$\vec{\mathbf{\Delta}} = (\vec{\Delta}_i)_i$ fluctuations of the i-th landmark Cartesian coordinates wrt. their averages



Unsupervised inference: $\mathcal{S} = \{\vec{\mathbf{\Delta}}^{(s)}\}_{s=1}^{S} \to P(\vec{\mathbf{\Delta}}|\boldsymbol{\theta})$

Maximum Entropy inference

The most probable $\mathcal{L}(\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta})$ compatible with $C_{\alpha\beta}$

Maximum Entropy inference

The most probable $\mathcal{L}(\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta})$ compatible with $C_{\alpha\beta}$

$$\mathcal{L}(\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta}) = \frac{1}{Z} e^{-H[\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta}]}$$

where H is the effective Hamiltonian:

$$\begin{aligned} H[\vec{\boldsymbol{\Delta}}|J] &= \frac{1}{2}\sum_{\alpha\beta} \Delta_\alpha \Delta_\beta J_{\alpha\beta} \\ H[\vec{\boldsymbol{\Delta}}|J^{(\mathrm{x})}, J^{(\mathrm{y})}, J^{(\mathrm{xy})}] &= \frac{1}{2}\sum_{i,j} \vec{\Delta}_i \cdot \begin{pmatrix} {J^{(\mathrm{x})}}_{ij} & {J^{(\mathrm{xy})}}_{ij} \\ {J^{(\mathrm{xy})}}_{ji} & {J^{(\mathrm{y})}}_{ij} \end{pmatrix} \cdot \vec{\Delta}_j \end{aligned}$$

Maximum Entropy inference

The most probable $\mathcal{L}(\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta})$ compatible with $C_{\alpha\beta}$

$$\mathcal{L}(\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta}) = \frac{1}{Z} e^{-H[\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta}]}$$

where H is the effective Hamiltonian:

$$\begin{aligned} H[\vec{\boldsymbol{\Delta}}|J] &= \frac{1}{2}\sum_{\alpha\beta} \Delta_\alpha \Delta_\beta J_{\alpha\beta} \\ H[\vec{\boldsymbol{\Delta}}|J^{(\mathrm{x})}, J^{(\mathrm{y})}, J^{(\mathrm{xy})}] &= \frac{1}{2}\sum_{i,j} \vec{\Delta}_i \cdot \begin{pmatrix} {J^{(\mathrm{x})}}_{ij} & {J^{(\mathrm{xy})}}_{ij} \\ {J^{(\mathrm{xy})}}_{ji} & {J^{(\mathrm{y})}}_{ij} \end{pmatrix} \cdot \vec{\Delta}_j \end{aligned}$$

the value of $\boldsymbol{\theta}$ is fixed by Maximum Likelihood $\sum_s \ln \mathcal{L}(\vec{\boldsymbol{\Delta}}^{(s)}|\boldsymbol{\theta})$

Maximum Entropy inference

The most probable $\mathcal{L}(\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta})$ compatible with $C_{\alpha\beta}$, $C_{\alpha\beta\kappa}$

Maximum Entropy inference

The most probable $\mathcal{L}(\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta})$ compatible with $C_{\alpha\beta}$, $C_{\alpha\beta\kappa}$

$$\mathcal{L}(\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta}) = \frac{1}{Z} e^{-H_2[\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta}]-H_3[\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta}]}$$

where H_2, H_3 are the effective Hamiltonians:

$$\begin{aligned} H_2[\vec{\boldsymbol{\Delta}}|Q] &= \frac{1}{2}\sum_{\mu\nu} \Delta_\mu \Delta_\nu J_{\mu\nu} \\ H_3[\vec{\boldsymbol{\Delta}}|Q] &= \frac{1}{6}\sum_{\mu\nu\kappa} \Delta_\mu \Delta_\nu \Delta_\kappa Q_{\mu\nu\kappa} \end{aligned}$$

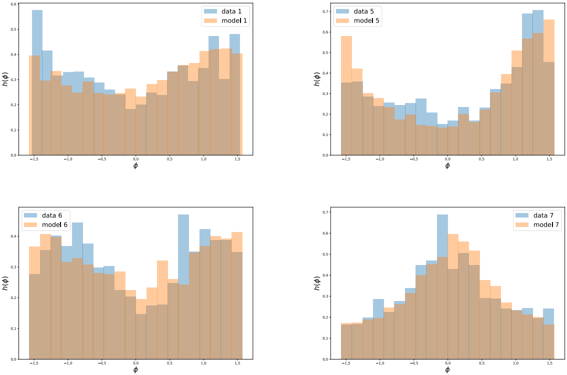
$\boldsymbol{\theta}$ are found by means on an efficient software based on perturbation expansions and the Wick's theorem [Monechi, I.-B. 2019?]

Maxent: Reproduction of nonlinear observables

Consider the observable $\phi_i^{(s)} = \mathrm{arctan}(\Delta_{i,\mathrm{y}}^{(s)}/\Delta_{i,\mathrm{x}}^{(s)})$

Maxent: Reproduction of nonlinear observables

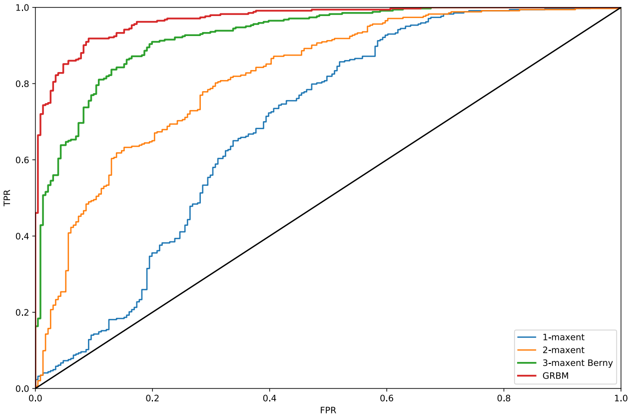
Consider the observable $\phi_i^{(s)} = \arctan(\Delta_{i,y}^{(s)}/\Delta_{i,x}^{(s)})$
The (harmonic) model approximately reproduces the experimental histogram of ϕ's



Facial preference II: inference of the database of sculpted vectors

$P(\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta})$ for classification of the subject's gender

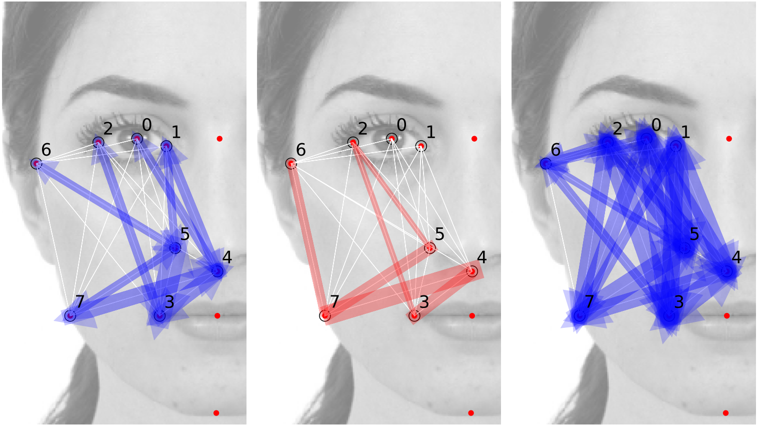
the female-score of $\boldsymbol{\Delta}$: $\mathsf{s}(\boldsymbol{\Delta}) = \ln P(\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta}_{\mathrm{female}}) - \ln P(\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta}_{\mathrm{male}})$



Facial preference II: inference of the database of sculpted vectors

MaxEnt inferred parameters $\boldsymbol{\theta}$ of $P(\vec{\boldsymbol{\Delta}}|\boldsymbol{\theta})$

$$\boldsymbol{\theta} = \{J_{ij}^{\|}, J_{ij}^{\perp}\}$$



Outline of inference results

Results regarding the inference of the experimental database $\mathcal{S} = \{\mathbf{x}^{(s)}\}_{s=1}^{S}$

Outline of inference results

Results regarding the inference of the experimental database $\mathcal{S} = \{\mathbf{x}^{(s)}\}_{s=1}^{S}$

- The distribution $\mathcal{L}(\cdot|\boldsymbol{\theta})$ is a faithful generatve model of the dataset

Outline of inference results

Results regarding the inference of the experimental database $\mathcal{S} = \{\mathbf{x}^{(s)}\}_{s=1}^{S}$

- The distribution $\mathcal{L}(\cdot|\boldsymbol{\theta})$ is a faithful generatve model of the dataset
  - It reproduces experimental non-linear functions of the data $\langle \mathcal{O} \rangle_{\mathrm{exp}} \simeq \langle \mathcal{O} \rangle_{\mathcal{L}}$
  - It performs well in a *classification* task (of the gender of the subject who sculped a given vector $\mathbf{x}^{(s)}$)

Outline of inference results

Results regarding the inference of the experimental database $\mathcal{S} = \{\mathbf{x}^{(s)}\}_{s=1}^{S}$

- The distribution $\mathcal{L}(\cdot|\boldsymbol{\theta})$ is a faithful generatve model of the dataset
  - It reproduces experimental non-linear functions of the data $\langle \mathcal{O} \rangle_{\rm exp} \simeq \langle \mathcal{O} \rangle_{\mathcal{L}}$
  - It performs well in a *classification* task
    (of the gender of the subject who sculped a given vector $\mathbf{x}^{(s)}$)
- Non-linear *effective interactions* are needed for a complete description of $\mathcal{S}$

Outline of inference results

Results regarding the inference of the experimental database $\mathcal{S} = \{\mathbf{x}^{(s)}\}_{s=1}^{S}$

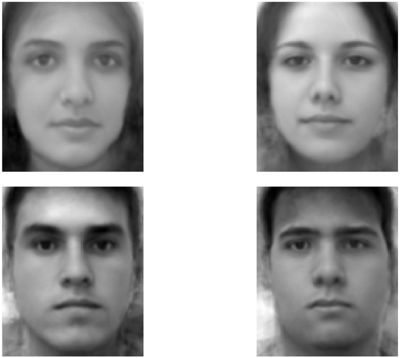
- The distribution $\mathcal{L}(\cdot|\boldsymbol{\theta})$ is a faithful generatve model of the dataset
  - It reproduces experimental non-linear functions of the data $\langle \mathcal{O} \rangle_{\rm exp} \simeq \langle \mathcal{O} \rangle_{\mathcal{L}}$
  - It performs well in a *classification* task (of the gender of the subject who sculped a given vector $\mathbf{x}^{(s)}$)
- Non-linear *effective interactions* are needed for a complete description of $\mathcal{S}$
- The matrix of effective interactions J provides relevant information (of cognitive origin) which goes beyond the raw, information present in the experimental correlations, C

Most informative face-space parametrisation

[Mariani, I.-B.]

Construction of a face space

Based on the separation of landmark/texture coordinates (inspired in the recently decorded neural coding for facial identification in the primate brain) [Chang+ 2018]



[Mariani, I.-B. 2019?]

Thank you !

miguel.berganza@roma1.infn.it

http://www.fis.unipr.it/home/miguel.berganza/