Image Statistics based on Diffeomorphic Matching

We consider a set of several images: I_i.

We want to define and compute:

$\hookrightarrow$ their mean image: M

$\hookrightarrow$ their characteristical ways of changing shape or colors

Mean image M:
supposed to *look like* each one of the images I_i.
$\hookrightarrow$ What does "*look like*" mean ?

$\Rightarrow$ We have to consider first the image matching problem.

Image matching

Let A and B be two images.

We think of them as positive real functions functions defined in a rectangular subset Ω of the plane $\mathbb{R}^2$.

$A, B : \mathbb{R}^2 \to \mathbb{R}$

We search for a deformation field $\mathbf{f}$ such that the warped image $A \circ \mathbf{f}$ resembles B.

$? \, \mathbf{f}, \;\; A \circ \mathbf{f} \simeq B$

The field $\mathbf{f}$ should be smooth enough and invertible:
$\Rightarrow$ take into account a regularizing term $R(\mathbf{f})$.

Examples:

$\hookrightarrow$ $\quad \|\mathbf{f} - Id\|_{\Omega}^{H^1}$

(where $\|a\|_{\Omega}^{H^1} = \int_{x \in \Omega} \|a(x)\|^2 + \|Da(x)\|^2 \, dx$)

$\hookrightarrow$ $\quad \|\mathbf{f} - Id\|_{\Omega}^{H^1} + \|\mathbf{f}^{-1} - Id\|_{\Omega}^{H^1}$.

Now, what does $A \simeq B$ mean ?

$\hookrightarrow$ criterium $C(A,B)$ which expresses the similarity between the two images A and B.

Example:

$\hookrightarrow$ $\|A-B\|_{\Omega}^{L^2} = \int_{x\in\Omega} (A(x)-B(x))^2\,dx$

$\hookrightarrow$ Local Cross-Correlation.

Local Cross-Correlation

Given a scale σ, the cross-correlation of two images A and B at point x is defined by:

$$CC(A, B, x) = \frac{v_{AB}(x)^2}{v_A(x)\, v_B(x)}$$

where $v_A(x)$ is the local spatial variance of A in a gaussian neighborhood of size σ centered on x, and $v_{AB}(x)$ the local covariance of A and B on the same neighborhood, i.e. we define:

$$g(x, y) = e^{\frac{\|x-y\|^2}{2\sigma^2}}$$

$$\mu(x) = \int_{y \in \Omega} g(x, y)\, dy$$

$$\bar{A}(x) = \frac{1}{\mu(x)} \int_{y \in \Omega} A(y)\, g(x, y)\, dy$$

$$v_A(x) = \epsilon + \frac{1}{\mu(x)} \int_{y \in \Omega} (A(y) - \bar{A}(x))^2 \, g(x,y) \, dy$$

$$v_{AB}(x) = \frac{1}{\mu(x)} \int_{\Omega} (A(y) - \bar{A}(x))(B(y) - \bar{B}(x)) \, g(x,y) \, dy$$

$$LCC(A,B) = \int_{x \in \Omega} CC(A,B,x) \, dx$$

The Image Matching Algorithm

Minimize with respect to the deformation field $\mathbf{f}$ (initialized to the identity) through a multi-scale gradient descent the following energy:

$$E(A,\ B,\ \mathbf{f})\ =\ LCC(A\circ\mathbf{f},\ B)\ +\ R(\mathbf{f})$$

Ref: O. Faugeras and G. Hermosillo, *Well-posedness of two non-rigid multimodal image registration methods*, Siam Journal of Applied Mathematics, 2004.

The mean of a set of images

An intuitive algorithm: find the mean

Introduce n diffeomorphisms $\mathbf{f}_i$ in order to warp an image A_i on the mean M.

$$? \; M, (\mathbf{f}_i)_{1 \leq i \leq n}, \qquad \text{min} \; \sum_i E(A_i \circ \mathbf{f}_i, \; M, \; \mathbf{f}_i)$$

Problem: a gradient descent with respect to an image (M) leads to bad results
$\hookrightarrow$ prevents the fields $\mathbf{f}_i$ from any evolution.

Another intuitive algorithm

Choose $M = \frac{1}{n} \sum_i A_i \circ \mathbf{f}_i$.

$$? \ \mathbf{f}_i, \qquad \min \ \sum_i E(A_i \circ f_i, \ \frac{1}{n} \sum_k A_k \circ \mathbf{f}_k, \ \mathbf{f}_i)$$

New problem: we have at each step for each i, $A_i \circ \mathbf{f}_i \ \simeq \ \frac{1}{n} \sum_i A_i \circ \mathbf{f}_i$
$\Rightarrow$ immediatly stuck in a local minimum.

The final word: eliminating the mean

$$? \, \mathbf{f}_i, \qquad \min \; \frac{1}{n-1} \sum_{i \neq j} LCC(A_i \circ \mathbf{f}_i, \; A_j \circ \mathbf{f}_j) \; + \; \sum_k R(\mathbf{f}_k)$$

At the end of the evolution:
$\hookrightarrow$ each $A_i \circ \mathbf{f}_i$ is supposed to look like each of the others
$\Rightarrow$ the mean is naturally computed as $M = \frac{1}{n} \sum_i A_i \circ \mathbf{f}_i$.

We add the condition $\sum_i \mathbf{f}_i = 0$ at each time step.

Example

Face database from Yale*.
$\hookrightarrow$ mean face out of photographs of ten different people with similar expressions, approximatively the same illumination and position conditions, and the same size (195 * 231 pixels).

*http://cvc.yale.edu/projects/yalefaces/yalefaces.html

The first 5 images.

The first 5 warped images.

The last 5 images A_i.

The last 5 warped images $A_i \circ \mathbf{f}_i$.

The mean of the previous ten faces.

Second order statistics of a set of images

Information about shape variations: in the diffeomorphisms $\mathbf{h}_i$
$\Rightarrow$ compute statistics on these warping fields.

Diffeomorphisms $\mathbf{f}_i$: functions from a subset Ω of the plane $\mathbb{R}^2$ to itself
$\Rightarrow$ correlation between two fields $\mathbf{a}$ and $\mathbf{b}$:

$$\langle \mathbf{a} \,|\, \mathbf{b} \rangle_{L^2(\Omega \to \mathbb{R}^2)} = \frac{1}{|\Omega|} \int_\Omega \mathbf{a}(x) \cdot \mathbf{b}(x) \, dx$$

Mean field $\bar{\mathbf{f}} = \frac{1}{n} \sum_i \mathbf{f}_i$ is 0
$\Rightarrow$ (shape-)correlation matrix:

$$SCM_{i,j} = \left\langle \mathbf{f}_i - \bar{\mathbf{f}} \,\middle|\, \mathbf{f}_j - \bar{\mathbf{f}} \right\rangle_{L^2(\Omega \to \mathbb{R}^2)} = \left\langle \mathbf{f}_i \,\middle|\, \mathbf{f}_j \right\rangle_{L^2(\Omega \to \mathbb{R}^2)}$$

$\hookrightarrow$ diagonalize the shape-correlation matrix SCM

$\hookrightarrow$ extract its eigenvalues σ_k and normalized eigenvectors $\mathbf{v}_k$

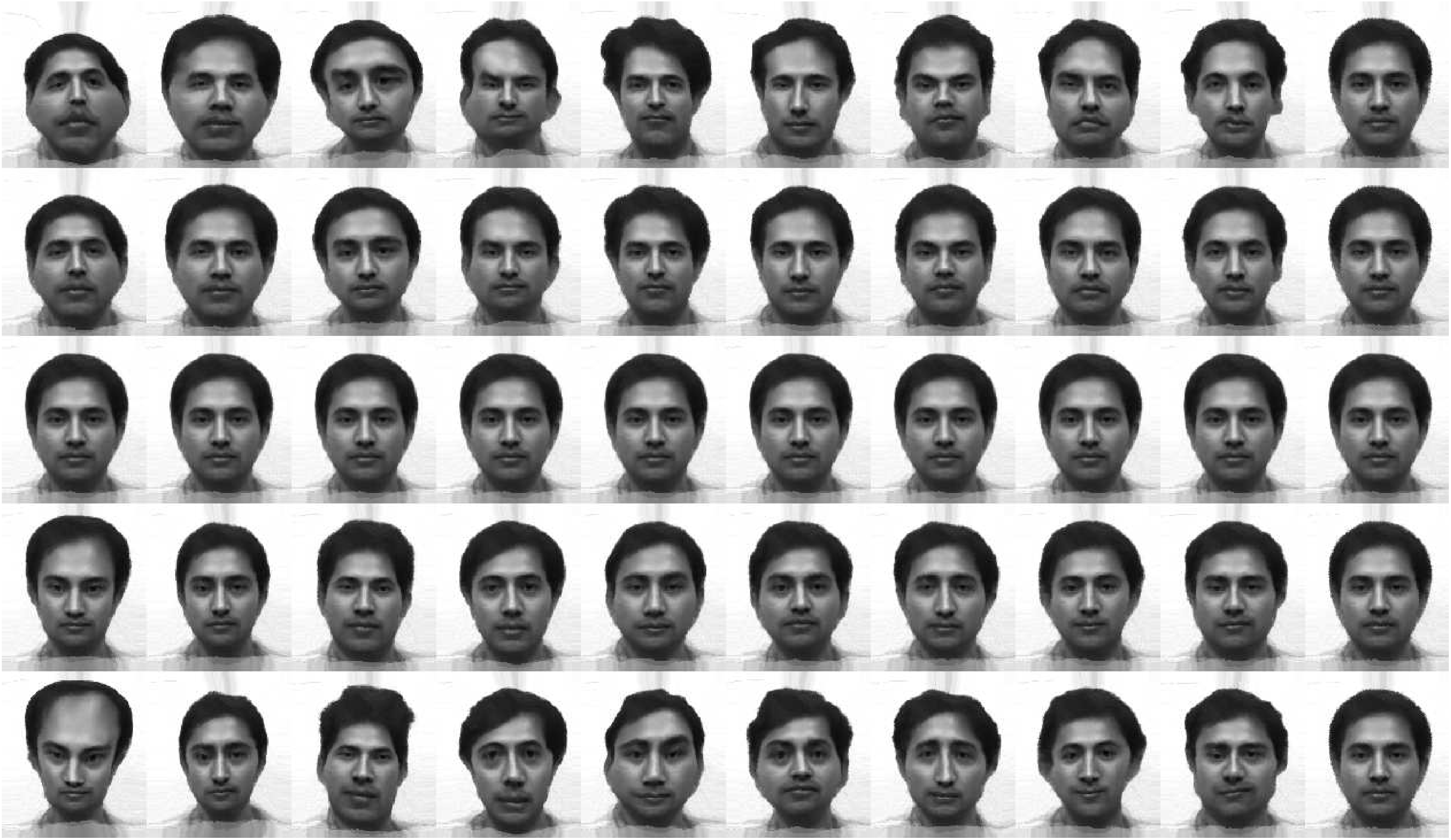
$\hookrightarrow$ modes of deformation: $\mathbf{m}_k = \sum_i (\mathbf{v}_k)_i \, \mathbf{f}_i$

In order to "draw" a mode: continuously apply a mode $\mathbf{m}_k$ to the mean image M with an amplitude α ($\in \mathbb{R}$) by computing the image $M \circ (Id + \alpha(\mathbf{m}_k - Id))$

$\Rightarrow$ make animations

Example

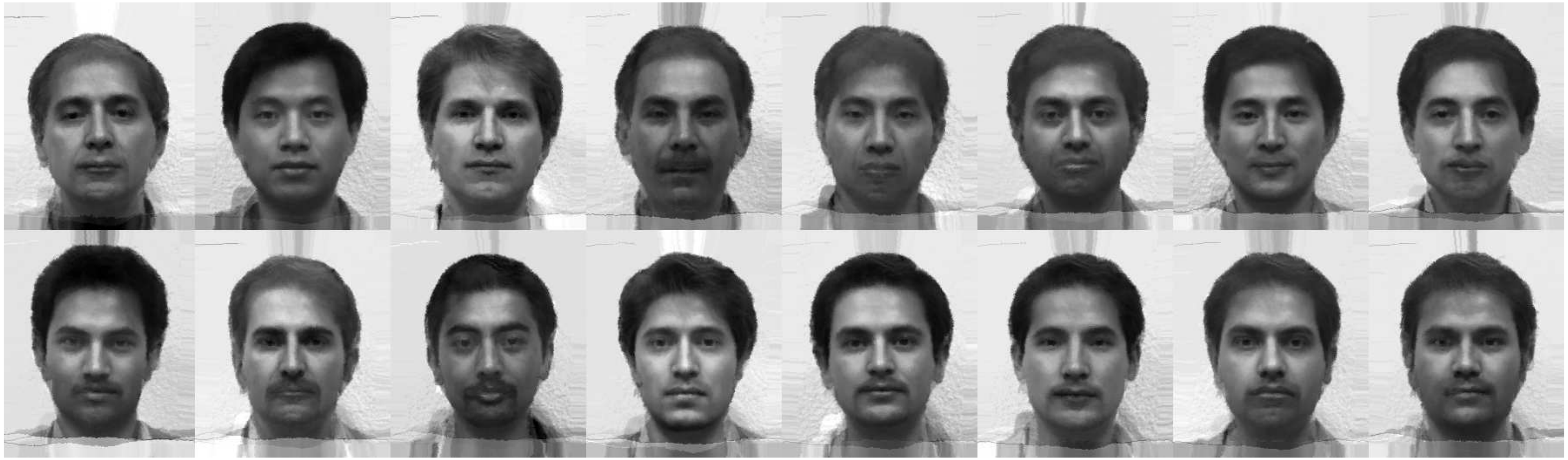
Each column represents a mode and is divided in five images: the mode is applied with five different amplitudes ($\alpha = \{2\sigma_k, \sigma_k, 0, -\sigma_k, -2\sigma_k\}$).



Intensity variations

Shape variations: statistics on diffeomorphisms $\mathbf{f}_i$.
Intensity variations: statistics on $I_i = A_i \circ \mathbf{f}_i - M$ for the $L^2(\mathbb{R}^2 \to \mathbb{R})$ scalar product.
$\hookrightarrow$ intensity-correlation matrix ICM with $ICM_{i,j} = \left\langle I_i \middle| I_j \right\rangle$.

Standard deviations of shapes and intensities: $\sigma_S^2 = \frac{1}{n} \sum_i \|\mathbf{f}_i\|^2$ and $\sigma_I^2 = \frac{1}{n} \sum_i \|I_i\|^2$
$\Rightarrow$ combined correlation matrix $CCM = 1/\sigma_S^2 SCM + 1/\sigma_I^2 ICM$
$\hookrightarrow$ proceed as before, compute and display principal modes of variations.



Each column represents a mode, applied to their mean image with amplitude $\alpha = \{\sigma_k, -\sigma_k\}$.

Classification: Expression Recognition

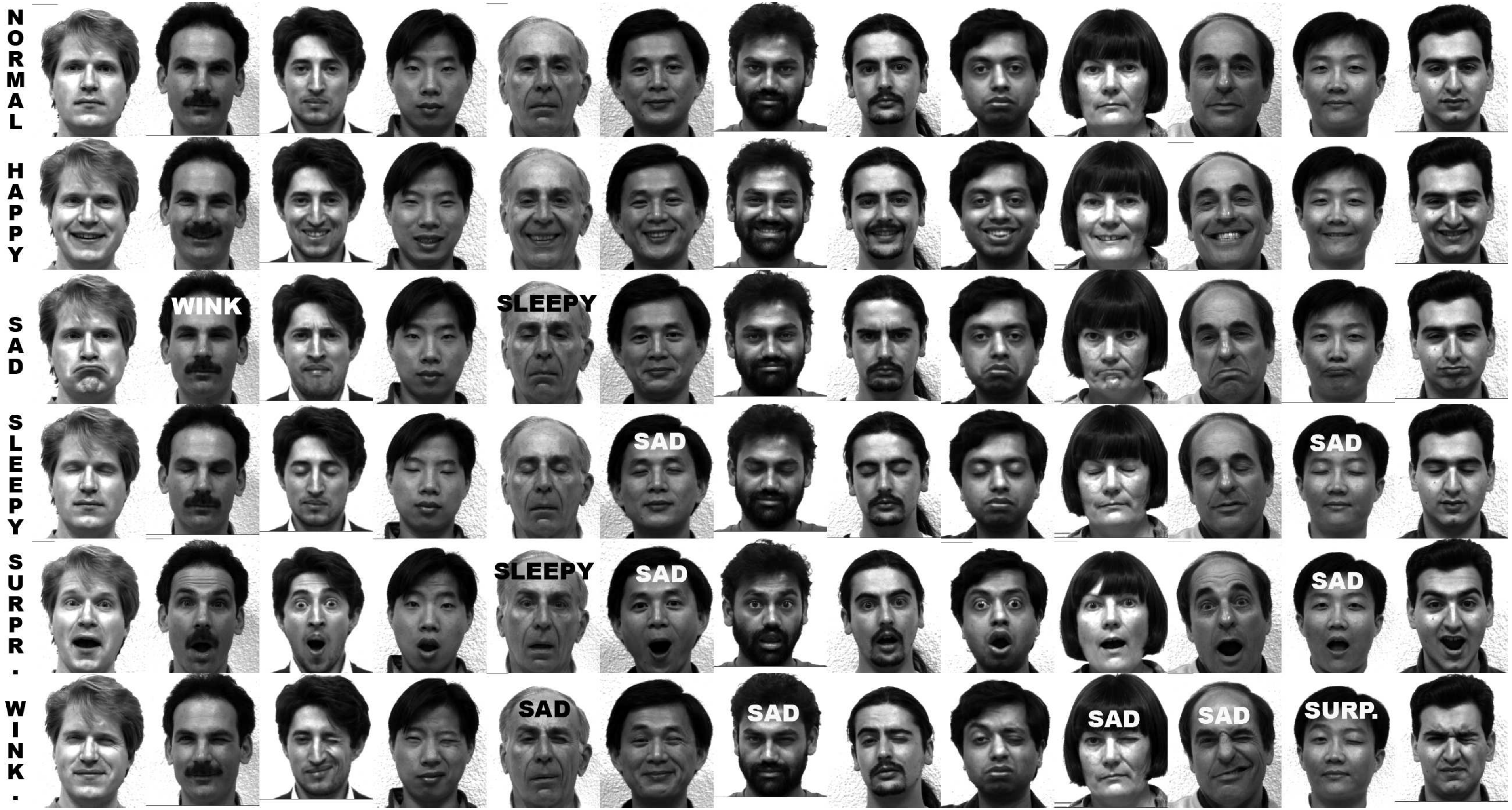
Associate with any new face its expression (out of happy, sad, sleepy, surprised and winking).

From the mean image

Match the mean image to the new image.
Use a Support Vector Machine with gaussian kernel on these deformations.
Cross-validation error: 24 upon 65 faces.

With knowledge of the face without expression

Match the image with normal expression to the image with unknown expression.
Align the computed deformation on the mean face of the whole set of images.
SVM on these aligned deformations: 12 errors on 65.


NORMAL
HAPPY
SAD
WINK
SLEEPY
SLEEPY
SAD
SAD
SURPR.
SLEEPY
SAD
SAD
WINK.
SAD
SAD
SAD
SAD
SURP.

Summary and Conclusions

$\hookrightarrow$ Definition and computation of first and second order statistics of a set of images with a diffeomorphic matching approach (without landmarks).
$\hookrightarrow$ use them in a classification task.

Methods not specific to faces and without any prior on the kind of images.