Image Statistics
based on
Diffeomorphic Matching
We consider a set of several images: $I_i$. We want to define and compute:

- their mean image: $M$
- their characteristical ways of changing shape or colors
Mean image $M$:
supposed to look like each one of the images $I_i$.
← What does “look like” mean?

⇒ 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 defined in a rectangular subset $\Omega$ of the plane $\mathbb{R}^2$.

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

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

$$? f, \ A \circ f \simeq B$$
The field $f$ should be smooth enough and invertible: 
$\Rightarrow$ take into account a regularizing term $R(f)$.

Examples:

$\leftarrow \|f - Id\|_{\Omega}^{H^1}$

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

$\leftarrow \|f - Id\|_{\Omega}^{H^1} + \|f^{-1} - Id\|_{\Omega}^{H^1}$. 
Now, what does $A \simeq B$ mean?

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

Example:

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

→ Local Cross-Correlation.
Local Cross-Correlation

Given a scale $\sigma$, 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 $\sigma$ 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 \( f \) (initialized to the identity) through a multi-scale gradient descent the following energy:

\[
E(A, B, f) = LCC(A \circ f, B) + R(f)
\]

The mean of a set of images

An intuitive algorithm: find the mean

Introduce \( n \) diffeomorphisms \( f_i \) in order to warp an image \( A_i \) on the mean \( M \).

\[
\begin{align*}
? M, (f_i)_{1 \leq i \leq n}, & \quad \text{min} \sum_i E(A_i \circ f_i, M, f_i)
\end{align*}
\]

Problem: a gradient descent with respect to an image \((M)\) leads to bad results
\(-\) prevents the fields \( f_i \) from any evolution.
Another intuitive algorithm

Choose \( M = \frac{1}{n} \sum_i A_i \circ f_i \).

? \( f_i \), \( \min \sum_i E(A_i \circ f_i, \frac{1}{n} \sum_k A_k \circ f_k, f_i) \)

New problem: we have at each step for each \( i \), \( A_i \circ f_i \approx \frac{1}{n} \sum_i A_i \circ f_i \)
\( \Rightarrow \) immediately stuck in a local minimum.
The final word: eliminating the mean

\[ \min_{f_i} \left\{ \frac{1}{n-1} \sum_{i \neq j} LCC(A_i \circ f_i, A_j \circ f_j) + \sum_k R(f_k) \right\} \]

At the end of the evolution:
← each \( A_i \circ f_i \) is supposed to look like each of the others
⇒ the mean is naturally computed as \( M = \frac{1}{n} \sum_i A_i \circ f_i \).

We add the condition \( \sum_i f_i = 0 \) at each time step.
Example

Face database from Yale*.

*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 $A_i \circ f_i$. 
The mean of the previous ten faces.
Second order statistics of a set of images

Information about shape variations: in the diffeomorphisms $h_i$ ⇒ compute statistics on these warping fields.

Diffeomorphisms $f_i$: functions from a subset $\Omega$ of the plane $\mathbb{R}^2$ to itself ⇒ correlation between two fields $a$ and $b$:

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

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

$$SCM_{i,j} = \langle f_i - \bar{f} \big| f_j - \bar{f} \rangle_{L^2(\Omega \to \mathbb{R}^2)} = \langle f_i \big| f_j \rangle_{L^2(\Omega \to \mathbb{R}^2)}$$
diagonalize the shape-correlation matrix $SCM$
extract its eigenvalues $\sigma_k$ and normalized eigenvectors $v_k$
modes of deformation: $m_k = \sum_i (v_k)_i f_i$

In order to “draw” a mode: continuously apply a mode $m_k$ to the mean image $M$ with an amplitude $\alpha (\in \mathbb{R})$ by computing the image $M \circ (Id + \alpha (m_k - Id))$
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 $f_i$.

Intensity variations: statistics on $I_i = A_i \circ f_i - M$ for the $L^2(\mathbb{R}^2 \to \mathbb{R})$ scalar product.

→ intensity-correlation matrix $ICM$ with $ICM_{i,j} = \langle I_i \mid I_j \rangle$.

Standard deviations of shapes and intensities: $\sigma^2_S = \frac{1}{n} \sum_i \| f_i \|^2$
and $\sigma^2_I = \frac{1}{n} \sum_i \| I_i \|^2$

⇒ combined correlation matrix $CCM = 1/\sigma^2_S SCM + 1/\sigma^2_I ICM$

← 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.
Summary and Conclusions

Definition and computation of first and second order statistics of a set of images with a diffeomorphic matching approach (without landmarks).

use them in a classification task.

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