Statistical Models and Classification

Joakim Anden, Joan Bruna, Stéphane Mallat, Laurent Sifre, Irène Waldspurger

IHES
Ecole Polytechnique
• Considerable variability in each class.
• A Euclidean norm does not measure signal «similarities». 
Texture Discrimination

- Textures define high-dimensional image classes.
  - Realizations of stationary processes \( X \) but typically not Gaussian, not Markovian and not characterized by second order moments.

same power spectrum

\[\begin{array}{cccc}
\text{same power spectrum} & & & \\
\text{same power spectrum} & & & \\
\end{array}\]
Signal Classification

- Very high dimensional space $N \geq 10^6$.
- Few training samples per class $P \ll N$.
- Signals do not belong to a low-dimensional manifold.

Space dimension
$10^6 \rightarrow \infty$

Lower dimensional manifold

$\|f - g\|$
$\|\Phi(f) - \Phi(g)\|$

Reduce variability due to
translations, transposition (audio)
rotations, scaling (images)
action of groups

Reduce structural variability

Unsupervised learning
- SIFT
- MFCC

Supervised learning
- GMM
- Bag of Features
- Eigenmaps
- Dictionary learning
- SVM
- PCA
Stable Translation Invariants

- **Invariance** to translations $x_c(t) = x(t - c)$

  $\forall c \in \mathbb{R}, \quad \Phi(x_c) = \Phi(x)$.

  *Null space information*

- **Metric stability** with deformations $x_\tau(t) = x(t - \tau(t))$

  $7 \ 9 \ 6 \ 6 \ 9$

  $8 \ 6 \ 3 \ 4 \ 8$

  small deformations of $x \implies$ small modifications of $\Phi(x)$

  $\forall \tau, \quad \|\Phi(x_\tau) - \Phi(x)\| \leq C \sup_{t} |\nabla \tau(t)| \|x\|.$

- **Preserve information**

  deformation size
Overview

• Invariance and stability: Fourier failure and wavelet stability

• Scattering transform invariants and deep neural networks

• Representation of stationary textures and multifractals

• Classification of image and textures
• Fourier transform \( \hat{x}(\omega) = \int x(t) e^{-i\omega t} dt \) invariance:

\[
\text{if } x_c(t) = x(t - c) \text{ then } |\hat{x}_c(\omega)| = |\hat{x}(\omega)|
\]

• Instabilities to small deformations \( x_\tau(t) = x(t - \tau(t)) \):

\[
| |\hat{x}_\tau(\omega)| - |\hat{x}(\omega)|| \text{ is big at high frequencies}
\]

\[
\tau(t) = \epsilon t
\]
Wavelet Transform

- Dilated wavelets: \( \psi_\lambda(t) = 2^{-jQ} \psi(2^{-jQ}t) \) with \( \lambda = 2^{-jQ} \).

\[ \parallel Wx \parallel^2 = \parallel x \star \phi \parallel^2 + \sum_{\lambda} \parallel x \star \psi_\lambda \parallel^2 = \parallel x \parallel^2. \]

- Wavelet transform: \( Wx(t) = \left\{ x \star \phi(t), x \star \psi_\lambda(t) \right\}_\lambda \)

- If \( |\hat{\phi}(\omega)|^2 + \sum_{\lambda} |\hat{\psi}_\lambda(\omega)|^2 = 1 \) then \( W \) is unitary:

\[ \parallel Wx \parallel^2 = \parallel x \star \phi \parallel^2 + \sum_{\lambda} \parallel x \star \psi_\lambda \parallel^2 = \parallel x \parallel^2. \]
• For images, dilated and rotated wavelets:

\[ \psi_\lambda(t) = 2^j \psi(2^j r t) \quad \text{with} \quad \lambda = 2^j r \]

• Wavelet transform: \( W x(t) = \left\{ x \ast \phi(t), x \ast \psi_\lambda(t) \right\}_\lambda \)

• If \( |\phi|^2 + \sum_\lambda |\hat{\psi}_\lambda|^2 = 1 \) then \( W \) is unitary.
Wavelet Stabilization

Window Fourier

Wavelet time-frequency

Time/Space averaging

\[ \left\{ |x \star \psi_\lambda(t)| \right\}_\lambda \]

Locally invariant to translations

and stable to deformations

MFSC (audio)

SIFT (images)

But loss of information.
Wavelet Stabilization

\[ \{ |x \ast \psi_\lambda(t)| \}_\lambda \]

Non-linearity is needed to have a non-zero invariant

Modulus has "optimality"

Locally invariant to translations and stable to deformations

MFSC (audio)
SIFT (images)

But loss of information.

Time/Space averaging 370ms window
Recovering Lost Information

- A modulus computes a lower frequency envelop

- The averaging $|x \star \psi_{\lambda_1}| \star \phi$ removes high frequencies:

- Wavelet transform: $\{ |x \star \psi_{\lambda_1}| \star \psi_{\lambda_2} \}_{\lambda_2}$

- Translation invariance by time averaging the amplitude:

$$\forall \lambda_1, \lambda_2, \quad | |x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi \quad : \text{stable to deformations}$$
For any path $p = (\lambda_1, \lambda_2, ..., \lambda_m)$ of order $m$

$$S[p] x(t) = | |x \star \psi_{\lambda_1} | \star \psi_{\lambda_2} | ... | \star \psi_{\lambda_m} | \star \phi(t)$$

depends upon normalized moments of order $2^m$
Amplitude Modulation

\[ S[\lambda_1] x(t) = |x \ast \psi_{\lambda_1}| \ast \phi(t) \]

\[ S[\lambda_1, \lambda_2] x(t) = | |x \ast \psi_{\lambda_1}| \ast \psi_{\lambda_2}| \ast \phi(t) \text{ for } \lambda_1 = 1977 \]

1977 Hz

512 ms window

18 Hz
Sounds with Same Spectrum

\( X: \) stationary process

\[
\log(\lambda_1) \quad \text{J. McDermott} \quad \quad \quad \quad |X \ast \psi_{\lambda_1}|(x)
\]

\[
\log(\lambda_1) \quad 2s \text{ window} \quad \quad \quad \quad S[\lambda_1]X(t) = |X \ast \psi_{\lambda_1}| \ast \phi(t)
\]

\[
\log(\lambda_2) \quad \quad \quad \quad S[\lambda_1, \lambda_2]X(t) = ||X \ast \psi_{\lambda_1} \ast \psi_{\lambda_2}| \ast \phi(t) \quad \text{for} \ \lambda_1 = \log(1122)
\]
Image Wavelet Scattering

Images

\[ f \]

\[ \hat{f} \]

\[ \omega_1 \]

\[ \omega_2 \]

Wavelet Scattering

\[ |f \ast \psi_{\lambda_1}| \ast \phi \]

\[ ||f \ast \psi_{\lambda_1}| \ast \psi_{\lambda_2}| \ast \phi \]

SIFT

window size = image size

Saturday, July 7, 2012
Textures with Same Spectrum

$X$: stationary process

Textures $X$

Fourier Power Spectrum

Wavelet Scattering

$|X \ast \psi_{\lambda_1} \ast \phi | = |X \ast \psi_{\lambda_1} \ast \psi_{\lambda_2} \ast \phi |$

window size = image size
• Iteration on $Ux = \{x \star \phi, |x \star \psi_{\lambda}|\}_\lambda$, contracting.

- Output at all layers: $\{S[p]x\}_{p \in \mathcal{P}}$.

MFSC and SIFT are 1st layer outputs: $S[\lambda_1]x$
For any path $p = (\lambda_1, \lambda_2, ..., \lambda_m)$ of order $m$

$$S[p]x(t) = |x \ast \psi_{\lambda_1}| \ast \psi_{\lambda_2}| \ast \psi_{\lambda_m}| \ast \phi(t)$$

$$\|Sx\|^2 = \sum_{p \in P} \|S[p]x\|^2$$

**Theorem:** For appropriate wavelets, a scattering is

- **contracting** $\|Sx - Sy\| \leq \|x - y\|$  
- **preserves energy** $\|Sx\|^2 = \|x\|^2$  
- **stable to deformations** $\|Sx - Sx_\tau\| \leq C \sup_t |\nabla \tau(t)| \|x\|$  
- **invariant to translation** when the size of $\phi$ goes to $\infty$.  

Saturday, July 7, 2012
Energy Conservation

\[ \|Ux\| = \|Wx\| = \|x\| \]

Proof: The modulus pushes the energy towards low frequencies.

- Fast decay across layers of \( \|U[p]x\| \Rightarrow \|Sx\| = \|x\| \)
- Reduced number of paths with non-negligible output.
- Computational complexity: \( O(N \log N) \).
Theorem For appropriate wavelets

\[ Ux = \left\{ x \ast \phi(t), |x \ast \psi_\lambda(t)| \right\}_\lambda \]

is invertible and the inverse is continuous.

Inverse scattering: \( x \ast \phi \quad \xrightarrow{\text{More precise if sparse}} \quad \text{More stable phase recovery if } \left\{ |x \ast \psi_\lambda(t)| \right\}_\lambda \text{ are sparse} \)

because fewer phase to compute

\[ S[\lambda_1]x \quad \rightarrow \quad |x \ast \psi_\lambda(t)| \quad \rightarrow \quad t \]

\[ S[\lambda_1, \lambda_2]x = |x \ast \psi_{\lambda_1} \ast \psi_{\lambda_2}| \ast \phi \]

Not exactly invertible because the last layer is missing.

Smaller information loss if sparse: sparse deconvolution.
Original audio signal $x$

Reconstruction from $Sx$ for a window of 3 s with $N$ samples $Q = 8$

From order 1 $S[\lambda_1]x : Q \log N$ coefficients

From order 2 $S[\lambda_1, \lambda_2]x : (Q \log N)^2 / 2$ coefficients
• If $X(t)$ is stationary then

$$U[p]X = | \cdots | X \star \psi_{\lambda_1} | \cdots | \star \psi_{\lambda_m} |$$

is stationary.

• Expected scattering:

$$\overline{S}X(p) = E(U[p]X)$$

depends on normalized moments of order $2^m$ of $X$.

• A windowed scattering

$$S[p]X(t) = U[p]X \star \phi(t)$$

is an unbiased estimator of $\overline{S}X(p) = E(U[p]x)$. 
Scattering White Noises

Constant Fourier power spectrum: $\hat{R}_X(\omega) = \sigma^2$.

Bernoulli $X(x)$

Gaussian White

$\int \hat{R}_X(\omega) |\hat{\psi}_{2j}(\omega)|^2 \, d\omega = \int_{2j}^{2^{j+1}} \overline{S}X(p(\omega))^2 \, d\omega$. 

$\overline{S}X(\lambda_1)$

$\overline{S}X(\lambda_1, \lambda_2)$

$\overline{S}X(\lambda_1, \lambda_2, \lambda_3)$

$\overline{S}X(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$
- Multifractal scaling:

\[ \overline{SX}(\lambda_1) \sim \lambda_1^{-\gamma_1} \]

\[ \frac{\overline{SX}(\lambda_1, \lambda_2)}{\overline{SX}(\lambda_1)} \sim (\lambda_2 \lambda_1^{-1})^{-\gamma_2} \]

<table>
<thead>
<tr>
<th>Process</th>
<th>( \gamma_1 )</th>
<th>( \gamma_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>White Gaussian</td>
<td>(-1/2)</td>
<td>(-1/2)</td>
</tr>
<tr>
<td>Fractional Brownian Noise ( B_H(t) )</td>
<td>( H )</td>
<td>(-1/2)</td>
</tr>
<tr>
<td>Mandelbrot cascade</td>
<td>( \gamma_1 )</td>
<td>0</td>
</tr>
<tr>
<td>NASDAQ:AAPL</td>
<td>( 2/3 )</td>
<td>(-0.15)</td>
</tr>
<tr>
<td>Dirac measure</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Poisson pp density ( \alpha )</td>
<td>0 if ( \lambda &lt; \alpha )</td>
<td>0 if ( \lambda_1 + \lambda_2 &lt; \alpha )</td>
</tr>
<tr>
<td></td>
<td>(-1/2) if ( \lambda \geq \alpha )</td>
<td>(-1/2) if ( \lambda_1 + \lambda_2 \geq \alpha )</td>
</tr>
</tbody>
</table>
the maximum entropy distribution is (Boltzman theorem):

$$p(x) = \frac{1}{Z} \exp\left(\sum_{p \in \mathcal{P}} \alpha_p U[p] x\right)$$

where $\alpha_p$ are Lagrange multipliers and $Z$ is defined by

$$\int p(x) \, dx = 1.$$

• Metropolis-Hasting algorithm samples the distribution, but computationally very expensive.

• Faster iterative algorithm with sparsity condition on $l^0$ norm.
• Estimation of $X(x)$ from $\log^2 N$ second order coefficients:
  - Original jackhammer
    - Synthesized
  - Original water
    - Synthesized
  - Original applause
    - Synthesized
Image Reconstruction

Original

Reconstructed
• Each class $X_k$ is represented by a scattering centroid $E(SX_k)$ and a space $V_k$ of principal variance directions (PCA).

Affine space model $A_k = E(SX_k) + V_k$. 
Digit Classification: MNIST

Wavelet Scattering

\[ x \]

\[ |x \ast \psi_{\lambda_1} \ast \phi(2^J n)| \]

\[ |x \ast \psi_{\lambda_1} \ast \psi_{\lambda_2} \ast \phi(2^J n)| \]

\[ 2^J = 8 : \text{window size} \]

\[ \text{cross-validated} \]
Digit Classification: MNIST

Classification Errors

<table>
<thead>
<tr>
<th>Training size</th>
<th>Conv. Net.</th>
<th>Scattering</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>7.2%</td>
<td>4.4%</td>
</tr>
<tr>
<td>5000</td>
<td>1.5%</td>
<td>1.0%</td>
</tr>
<tr>
<td>20000</td>
<td>0.8%</td>
<td>0.6%</td>
</tr>
<tr>
<td>60000</td>
<td>0.5%</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

LeCun et. al.
Classification of Textures

CUREt database
61 classes

Rotations and illumination variations.
Classification of Textures

$X$

$||X \ast \psi_{\lambda_1}|| \ast \phi$

$||X \ast \psi_{\lambda_1} \ast \psi_{\lambda_2}|| \ast \phi$

window size = image size

cross-validated
Classification of Textures

CUREt database
61 classes

Rotations and illumination variations.

Classification Errors

<table>
<thead>
<tr>
<th>Training per class</th>
<th>Fourier Spectr.</th>
<th>Markov Field</th>
<th>Scattering</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>2.15%</td>
<td>2.46%</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

Varma & Zisserman
Rotation and Affine Invariance

- Scatterings along translation, rotation and affine groups:

\[ x \xrightarrow{\text{Translat. Invar.}} \xrightarrow{\text{Rotation Invar.}} \xrightarrow{\text{Affine Invar.}} Sx \]

UIUC database:

Classification Errors

<table>
<thead>
<tr>
<th>Training</th>
<th>Translation</th>
<th>Transl + Rotation</th>
<th>Affine</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>15 %</td>
<td>3%</td>
<td>1%</td>
</tr>
</tbody>
</table>
Conclusion

• *Stable, informative invariants* to think classification.

• Deformation stability need wavelets

• Sparsity is needed to preserve information in invariants

• Stochastic models are not understood.

• Unsupervised invariant learning is needed for complex classes

• Papers/softwares:  [www.cmap.polytechnique.fr/scattering](http://www.cmap.polytechnique.fr/scattering)