Independent Components of Natural Scenes are Edge Filters

Craig Calkins
February 7, 2012
What are we doing?

• Look at a linear model for images: $z = Wx$

• Attempt to learn independent filters that can be used to decompose a natural scene

• Compare the filters to other methods

• Discuss implications
Image Processing Background

• Can you tell what is in this image?
Image Processing Background

• The previous zebra image was a composite of these two
• It was the phase of the zebra combined with the amplitude of the cheetah
Image Processing Background

- Frequency – how quickly the intensity of the pixels change
- Phase – determines how the different frequencies will add up
Why independent filters?

• Simple cells in V1 respond to edge filters

• Low redundancy

• Has been done using sparse coding (Olshausen & Field 1996)
Image Representation

- Image – $x$
- Causes – $s$
- Basis Functions – $A$
- Filters – $W$
- Causes – $u$

$x = As$

$u = WAs = PSs$
Find Decorrelating $W$

\[ \langle uu^T \rangle = \text{Diagonal matrix} \]

\[ W^T W = \langle xx^T \rangle^{-1} \]

- Orthogonal Solution – PCA
- Symmetrical Solution – ZCA
- Independent Solution – ICA
PCA

- Global in space
- Local in frequency

\[ \text{EDE}^{-1} = \langle \mathbf{x} \mathbf{x}^T \rangle \]

\[ \mathbf{W} = \mathbf{D}^{-\frac{1}{2}} \mathbf{E}^T \]

\( \mathbf{E} \) is the eigenvectors of the covariance matrix
\( \mathbf{D} \) is a diagonal matrix of eigenvalues

\( \mathbf{W} \) represents the filters
\( \mathbf{E} \) would correspond to the bases
ZCA

- Local in space
- Global in frequency

\[ \mathbf{WW}^T = \langle \mathbf{x} \mathbf{x}^T \rangle^{-1} \]

\[ \mathbf{W} = \mathbf{W}^T \]

\[ \mathbf{W} = \langle \mathbf{x} \mathbf{x}^T \rangle^{-1/2} \]
ICA

- Semi-local filter
- Responds to both frequency and phase
<table>
<thead>
<tr>
<th></th>
<th>ICA</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>11</td>
<td>89</td>
<td>89</td>
</tr>
<tr>
<td>15</td>
<td>109</td>
<td>109</td>
</tr>
<tr>
<td>22</td>
<td>144</td>
<td>144</td>
</tr>
</tbody>
</table>
ICA Approaches

• Maximize non-gaussianity

• **Maximize entropy:**
  \[ H(A,B) = H(A) + H(B) - I(A,B) \]

• Use *infomax*, FastICA, JADE, Soft-Loss
Algorithm

- $H(y)$ is the joint entropy
- $H(y) = -\sum p(y_i) \ln(p(y_i))$
- $y = g(u)$ where $g$ is a non-linear function

$$\Delta W \propto \frac{\partial H(y)}{\partial W}$$

$$\Delta W \propto [W^T]^{-1} + \hat{y}x^T$$

$$\hat{y} = \frac{\partial}{\partial u} \ln \frac{\partial y}{\partial u}$$
Natural Gradient

- No longer dependent on a matrix inverse!

\[ \Delta W \propto (I + \hat{y}u^T)W \]
Difficulties of ICA

- May not be a solution
- Given algorithm may not find solution even if it exists, depends on choice of non-linearity $g$, and other not well understood reasons
Training

- 4 grayscale natural images
- Training set of 17,595 12x12 samples
- Update weights every 50 samples
- 30 sweeps of data
- 2 hours training time
Results

• High kurtosis

• Near independence

• Array of edge filters
Kurtosis

- Measure of the peakiness of a distribution
- High kurtosis means more sparseness
- ICA has higher kurtosis than PCA or ZCA

\[ K_i = \frac{\langle (u_i - \langle u_i \rangle)^4 \rangle}{\langle u_i^2 - \langle u_i^2 \rangle \rangle^2} \]
Independence

- Verified independence with pdf contour plots
- ICA is more independent than PCA or ZCA
Filters

- Filters vary both in orientation and frequency
- Many are similar to edge filters
- Similar to simple cell receptive fields in V1
Questions?