Independent Components of Natural Scenes are Edge Filters

Craig Calkins

February 7, 2012

What are we doing?

- Look at a linear model for images: $\mathbf{z} = \mathbf{W}\mathbf{x}$
- 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?

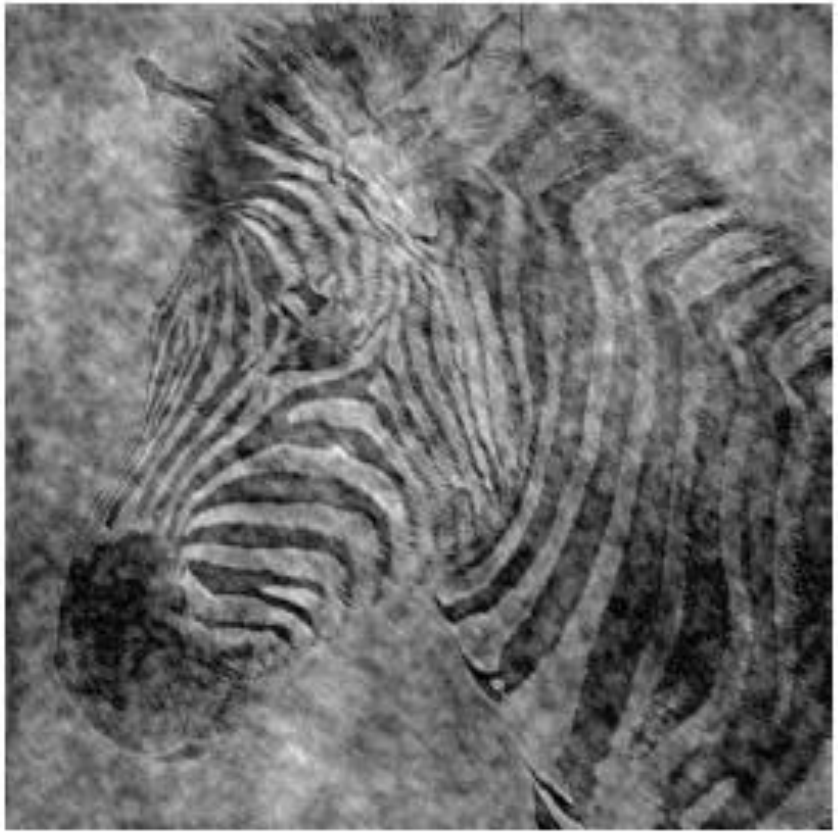


Image Processing Background

- 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

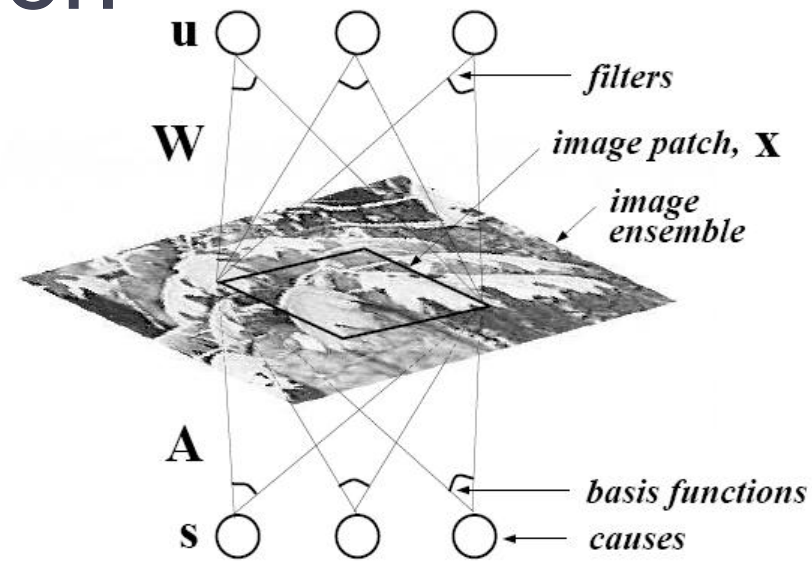
- Frequency – how quickly the intensity of the pixels change
- Phase – determines how the different frequencies will add up

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**



$$\mathbf{x} = \mathbf{As}$$

$$\mathbf{u} = \mathbf{Wx}$$

$$\mathbf{u} = \mathbf{WAs} = \mathbf{PSs}$$

Find Decorrelating W

$$\langle \mathbf{u}\mathbf{u}^T \rangle = \text{Diagonal matrix}$$

$$\mathbf{W}^T\mathbf{W} = \langle \mathbf{x}\mathbf{x}^T \rangle^{-1}$$

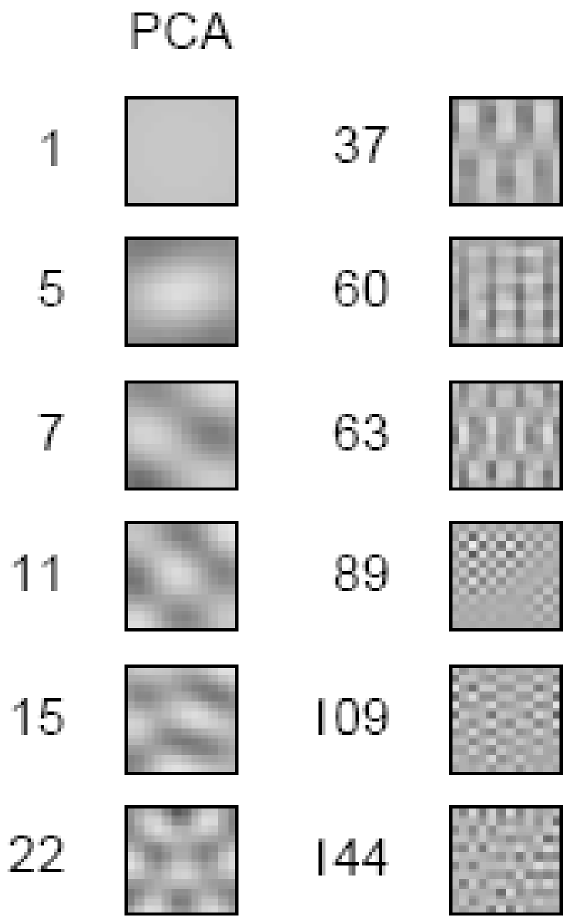
- Orthogonal Solution – PCA
- Symmetrical Solution – ZCA
- Independent Solution – ICA

PCA

- Global in space
- Local in frequency

- $\mathbf{EDE}^{-1} = \langle \mathbf{xx}^T \rangle$
- $\mathbf{W} = \mathbf{D}^{-1/2}\mathbf{E}^T$
- **E** is the eigenvectors of the covariance matrix
- **D** is a diagonal matrix of eigenvalues

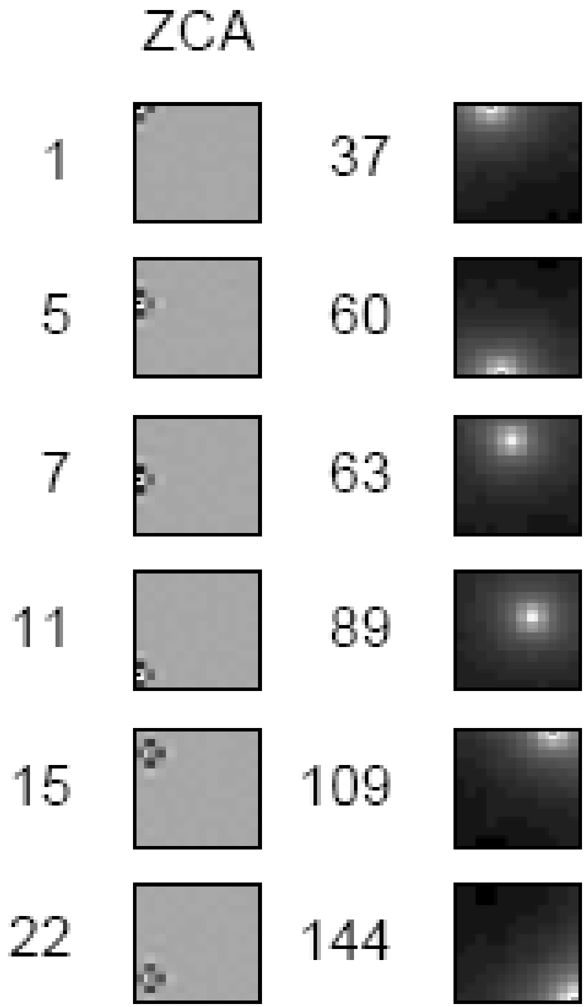
- **W** represents the filters
- **E** would correspond to the bases



ZCA

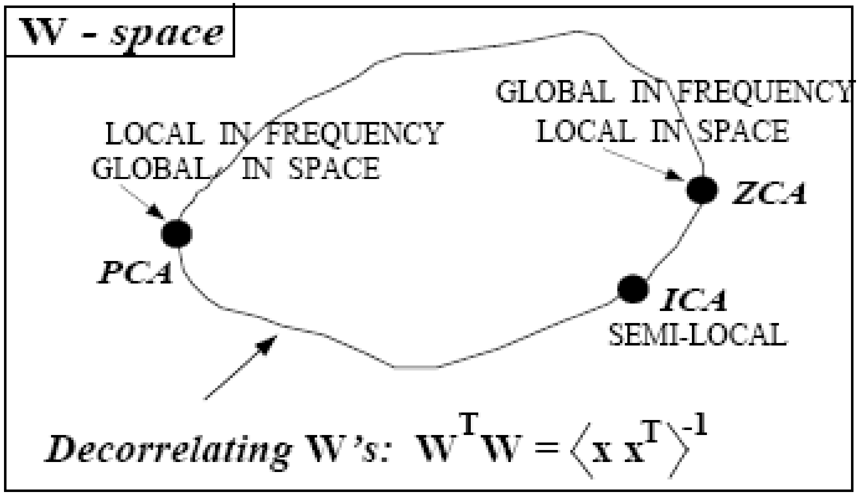
- Local in space
- Global in frequency

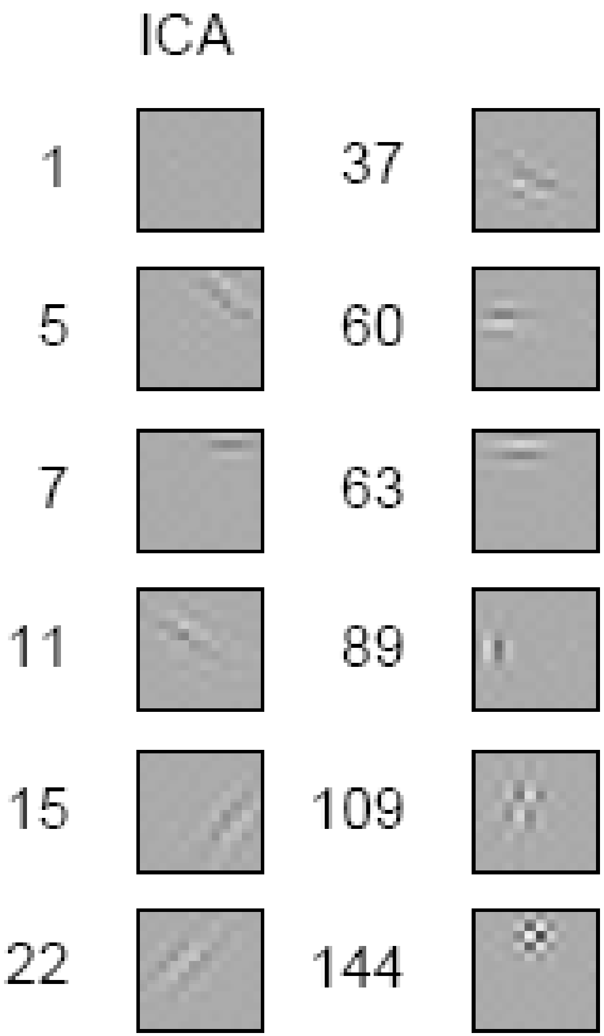
- $\mathbf{W}\mathbf{W}^T = \langle \mathbf{x}\mathbf{x}^T \rangle^{-1}$
- $\mathbf{W} = \mathbf{W}^T$
- $\mathbf{W} = \langle \mathbf{x}\mathbf{x}^T \rangle^{-1/2}$



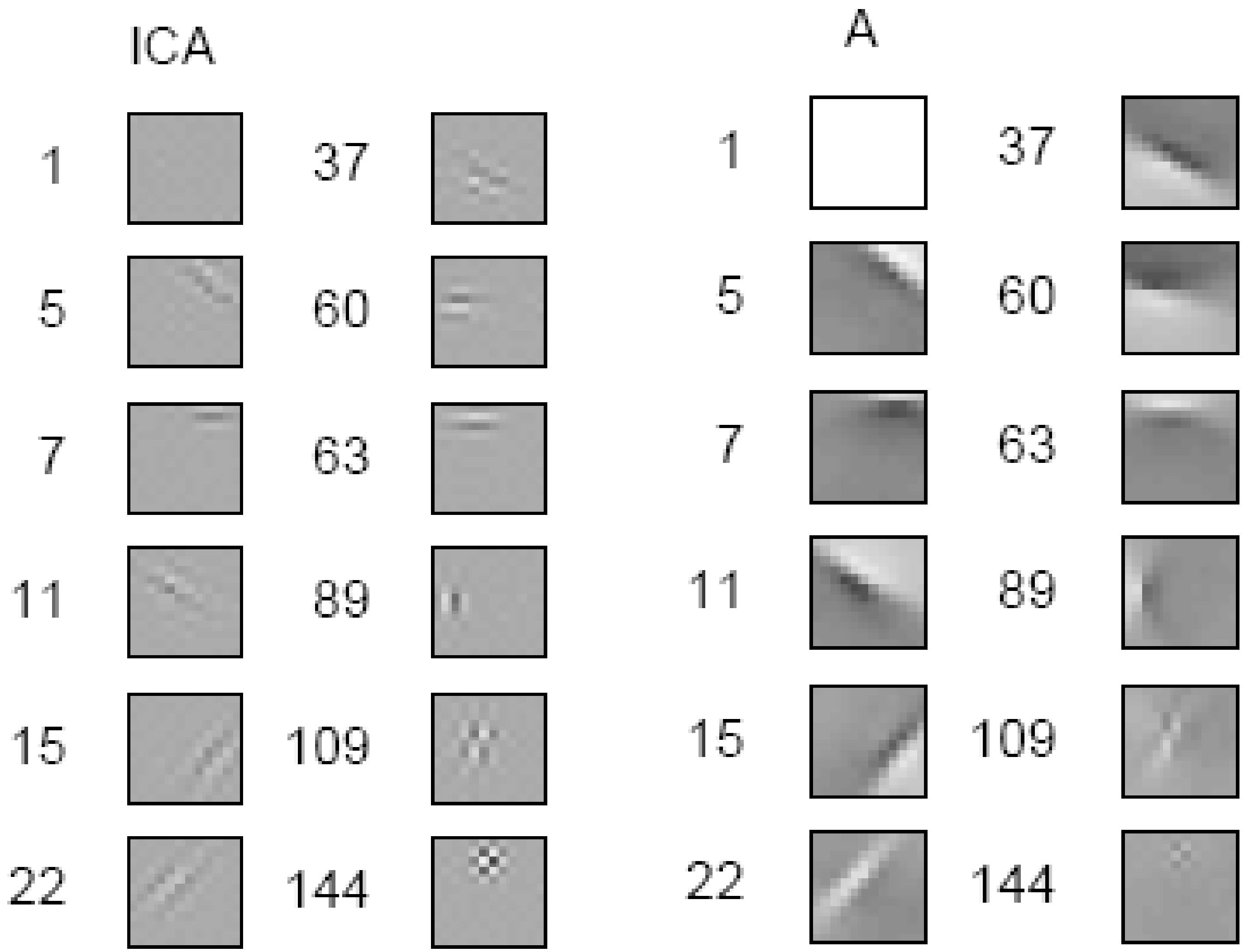
ICA

- Semi-local filter
- Responds to both frequency and phase





ICA


ICA
A
1
37
5
60
7
63
11
89
15
109
22
144
1
37
5
60
7
63
11
89
15
109
22
144

ICA Approaches

- Maximize non-gaussianity
- **Maximize entropy:**

$$H(\mathbf{A},\mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - I(\mathbf{A},\mathbf{B})$$

- 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

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

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

$$\hat{\boldsymbol{y}} = \frac{\partial}{\partial \boldsymbol{u}} \ln \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{u}}$$

Natural Gradient

- No longer dependent on a matrix inverse!

$$\Delta \boldsymbol{W} \propto (\boldsymbol{I} + \hat{\boldsymbol{y}}\boldsymbol{u}^T)\boldsymbol{W}$$

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



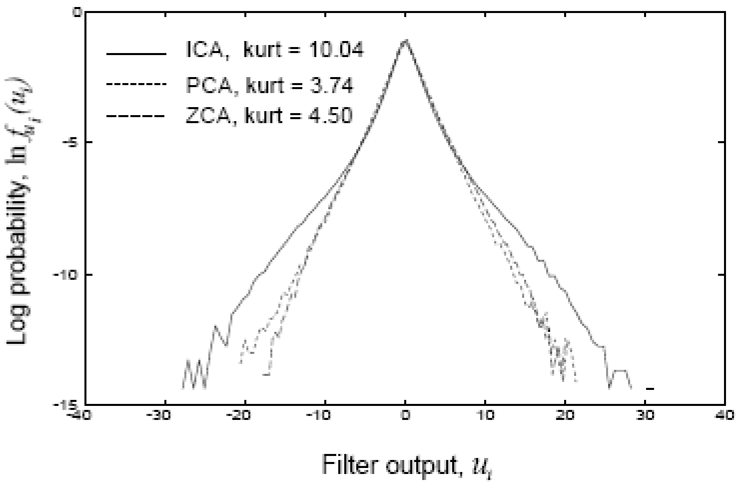

Results

- High kurtosis
- Near independence
- Array of edge filters

Kurtosis

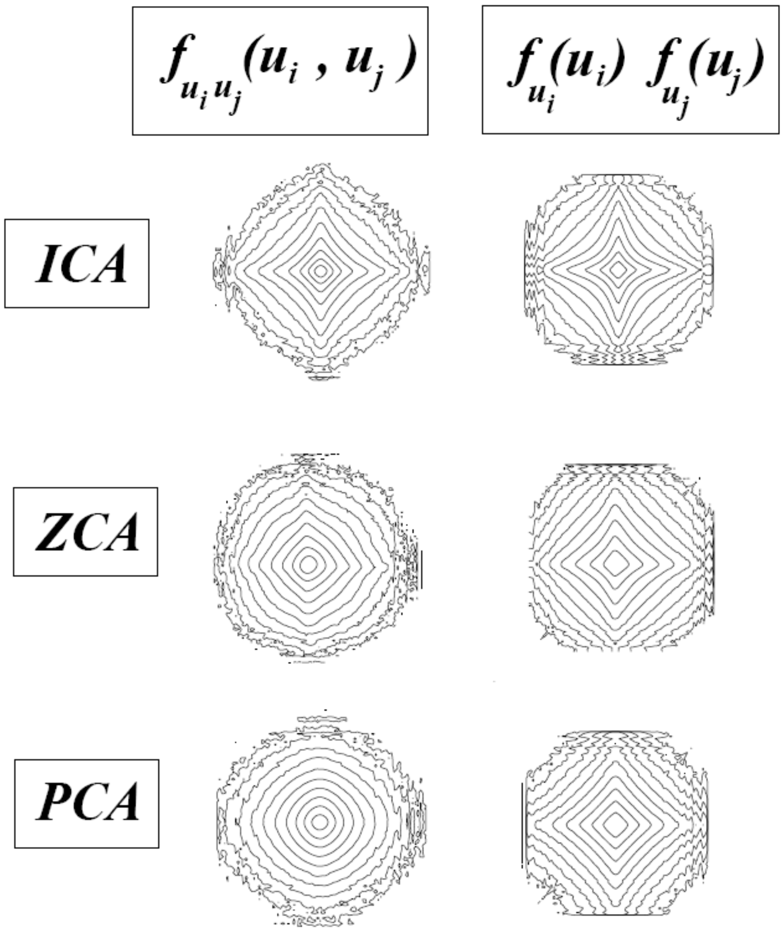
- Measure of the peakiness of a distribution
- High kurtosis means more sparseness
- ICA has higher kurtosis than PCA or ZCA

$$K_i = \frac{\langle (u_i - \langle u_i \rangle)^4 \rangle}{\langle u_i{}^2 - \langle u_i{}^2 \rangle \rangle^2}$$



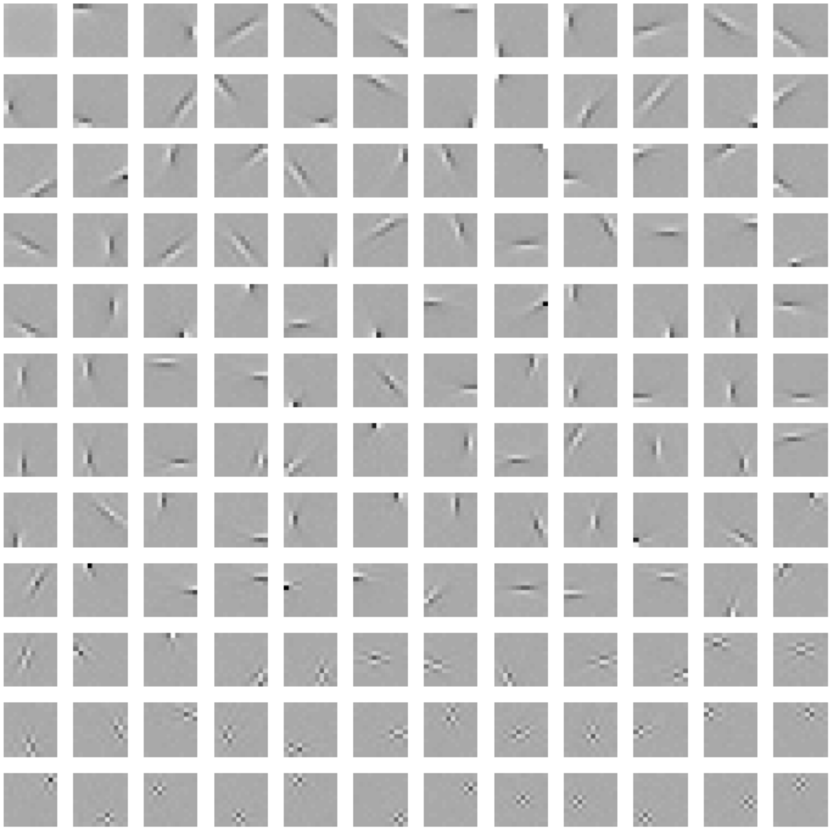
Independence

- 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?