Learning higher-order structures in natural images

Yan Karklin and Michael Lewicki

presented by David Ross
Feb 3, 2005

Efficient Coding

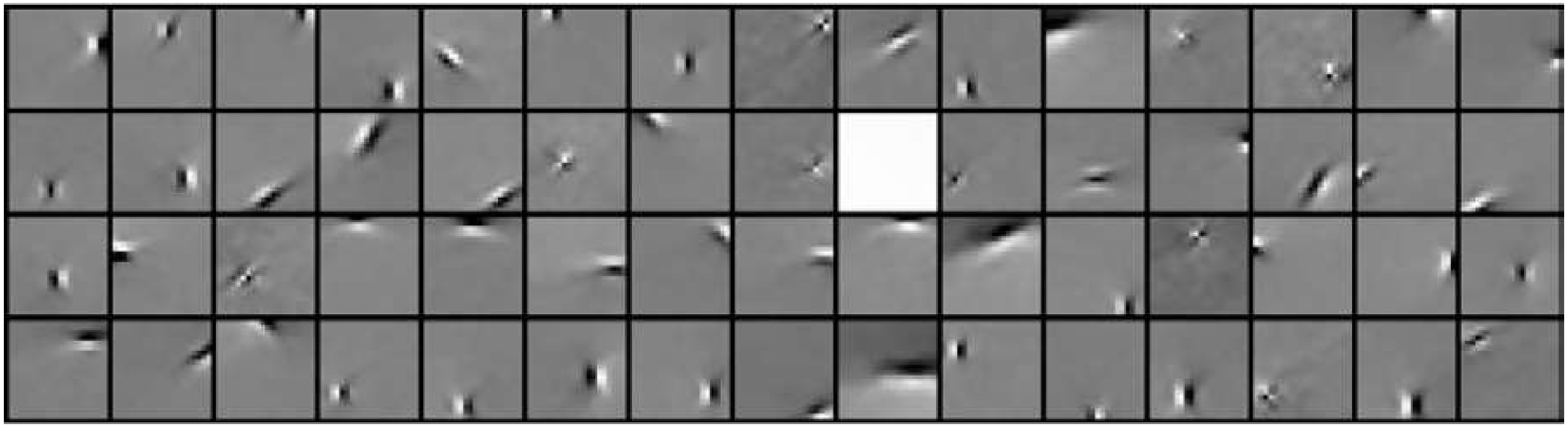
- efficient coding + linear models
  - produce Gabor-like filters
  - good at explaining early visual processing
  - only limited visual structure can be represented
- goal: learn higher-order image structure (e.g. object location, surface texture)
- approach: non-linear efficient codes, via a hierarchical model

ICA

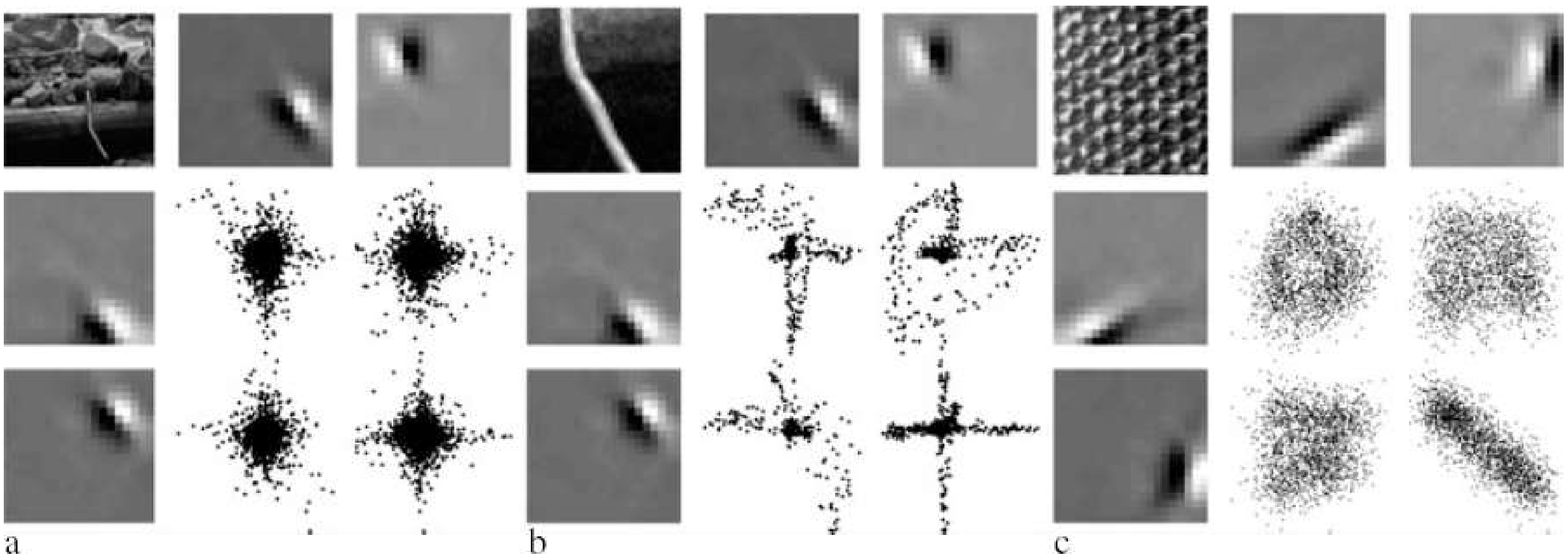
- Reduce redundancy in coefficients by making them independent
- square, noise-free: $x = W^{-1} u$
- probability model

$$p(x|W) = p(u)|W| \qquad p(u) = \prod_i p(u_i)$$

ICA basis functions



Coefficients are not independent



- why?
  - limitation of a linear model
  - only looking at statistical relationships among pixels

Hierarchical Model 1

- for each patch, the u's (coefficients) were not generated independently
- instead, variance of u's shows statistical regularities
  - by modeling these with efficient codes, hope to capture higher-order image structure
- 1st step: non-independent priors on the u's

$$p(\boldsymbol{u}|\lambda) = \exp\left(-\sum_i \left|\frac{u_i}{\lambda_i}\right|^q\right)$$

Hierarchical Model 2

- 2nd step: model variation in λ's between patches by learning an efficient code for it
  - v is code for λ's using following relationship

$$\lambda = \exp\left(B\,\boldsymbol{v}\right)$$

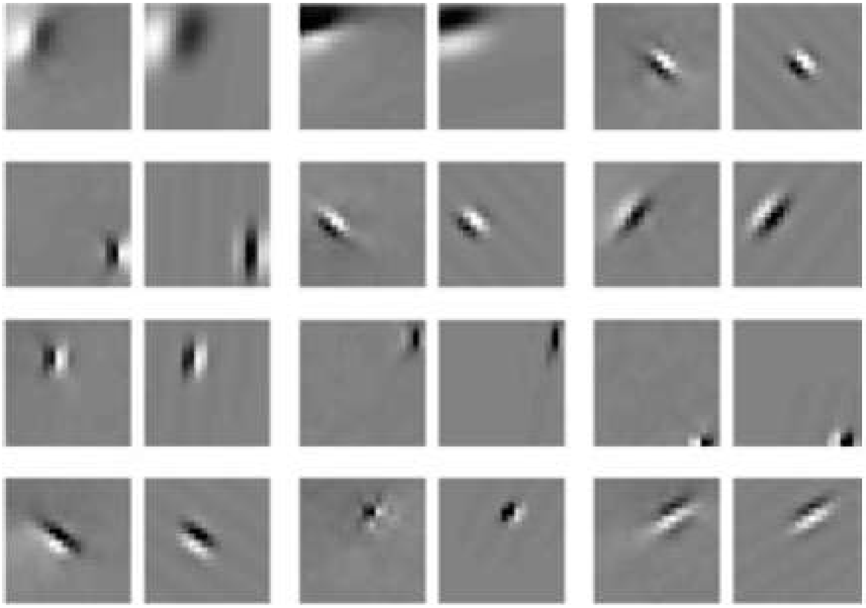
  - put sparse, independent prior on v's (or recurse)
- so log(λ) coded as a linear combination of some variance basis vectors B
- when v=0, λ=1 and we get ICA

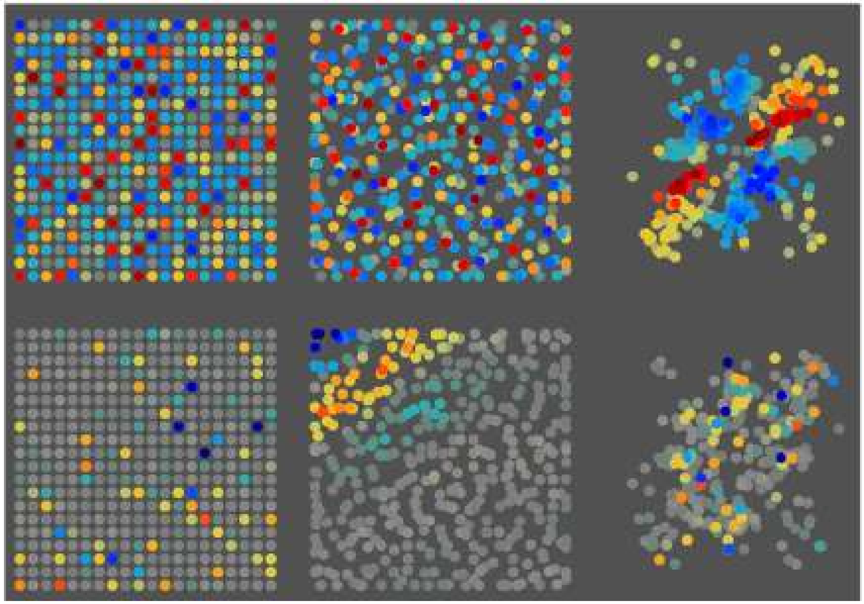
Inference & Learning

- they assume fixed image filters (W), as learned by ICA
- think of $u = Wx$ as the input
- get best v by maximizing $p(u|B,v)p(v)$ via gradient ascent
- solve for optimal* B, given u,v & Gaussian prior via gradient ascent
- "makes the assumption that the optimal solution for A is largely independent of the value of B"

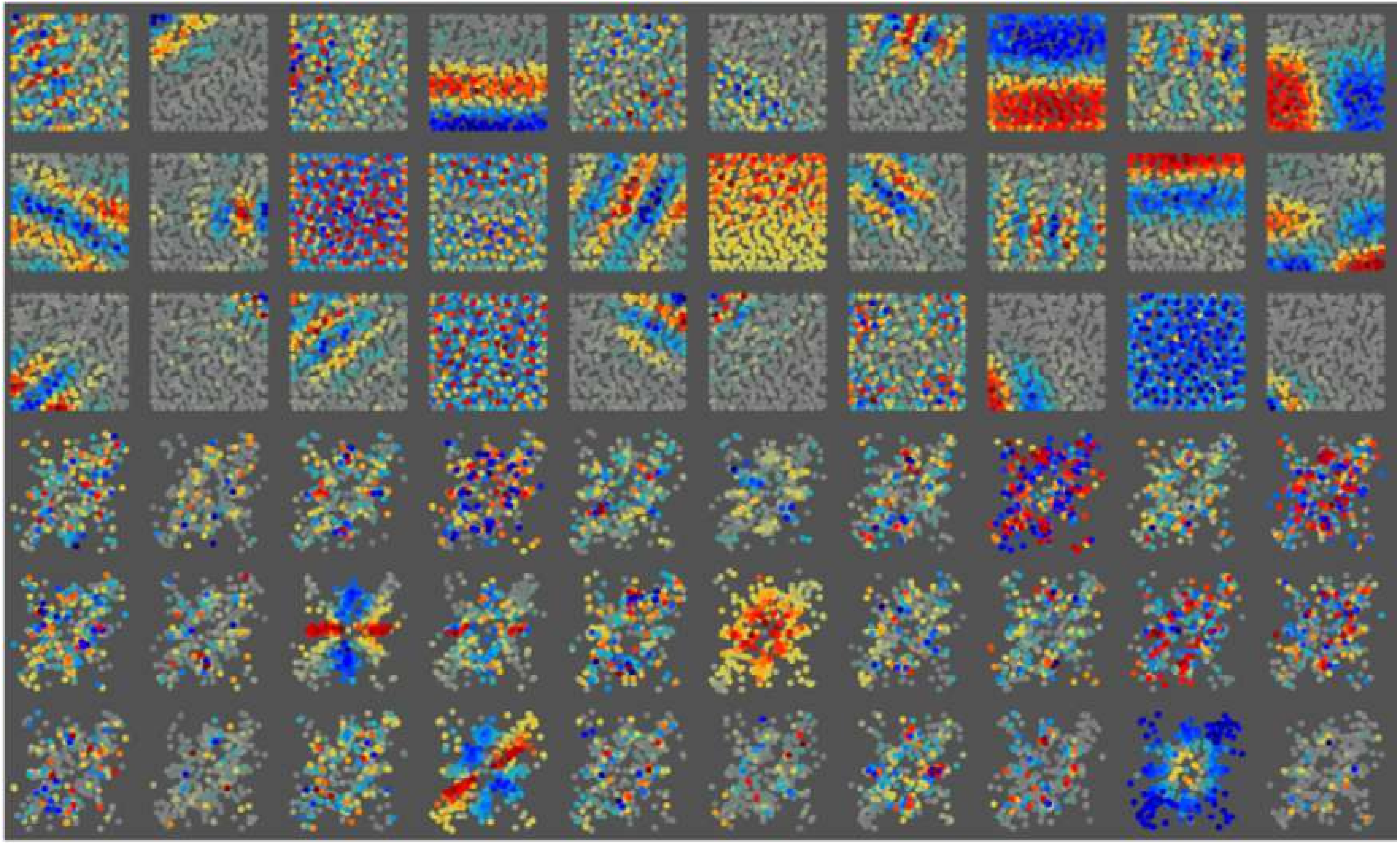
Visualize variance basis vectors

- hypothesis: v codes for higher-order structure

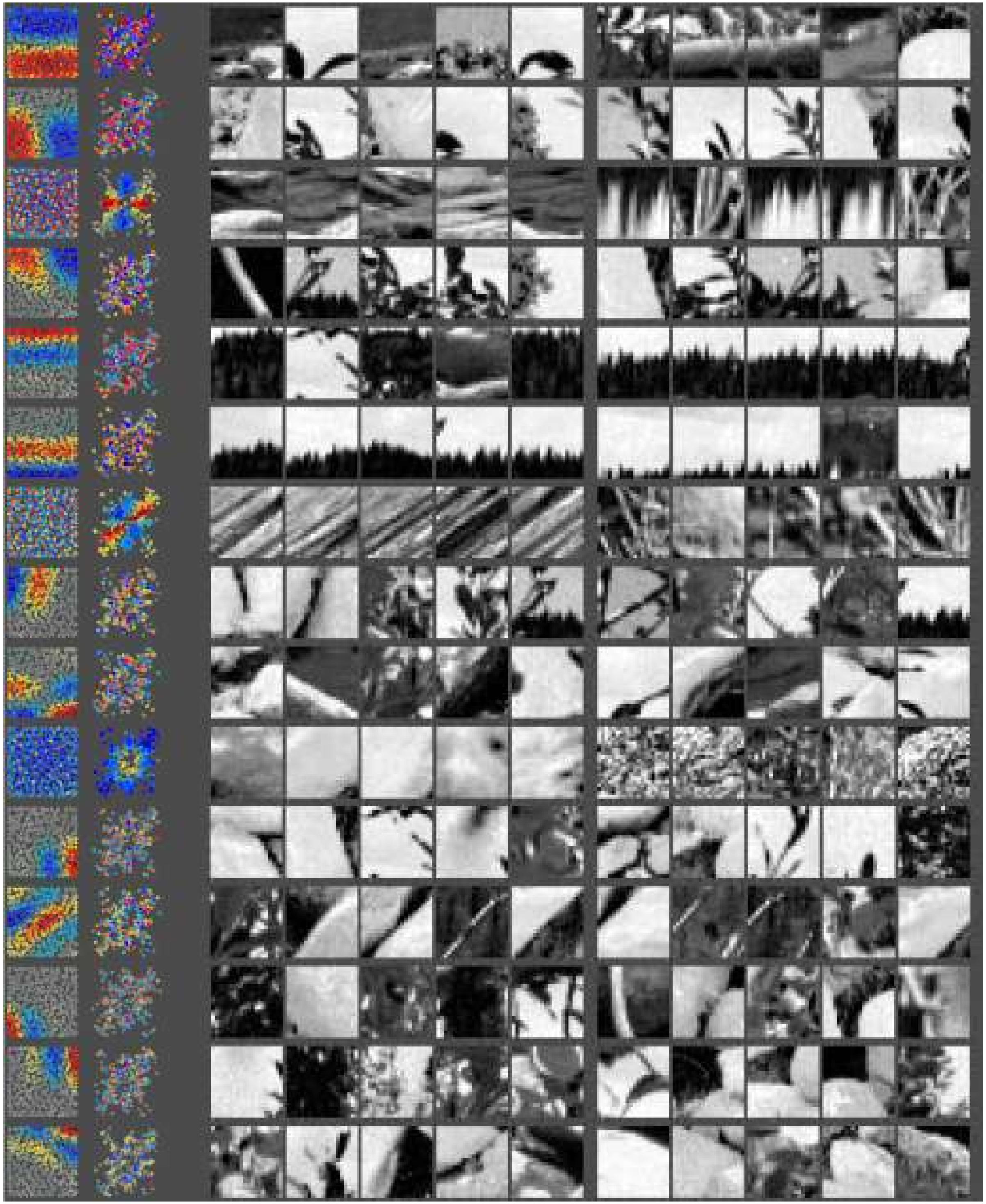




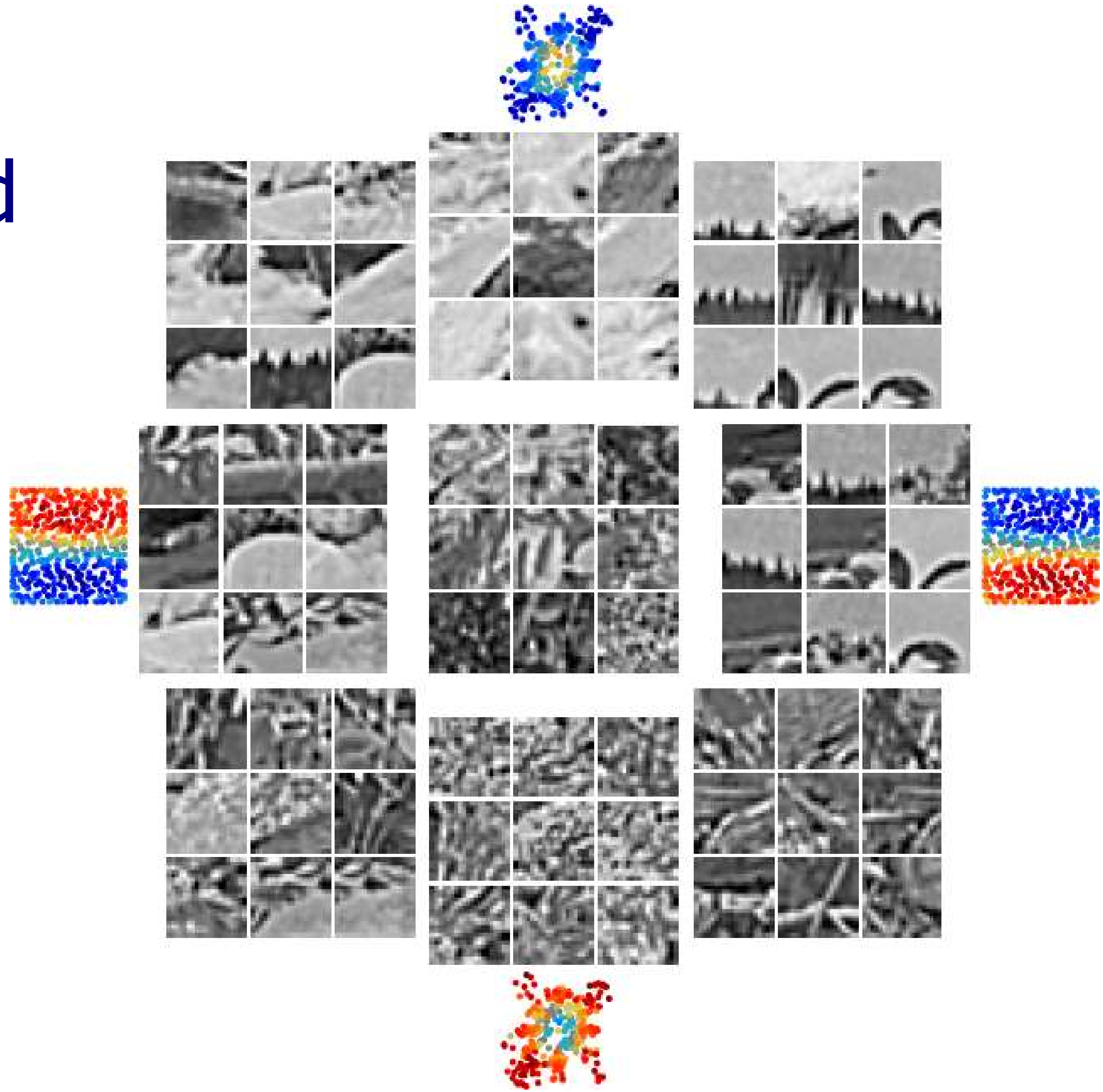
Contrast, Orientation, Frequency



Examples with high activation



v forms a distributed code



Coefficients (u's) don't capture higher-order structure

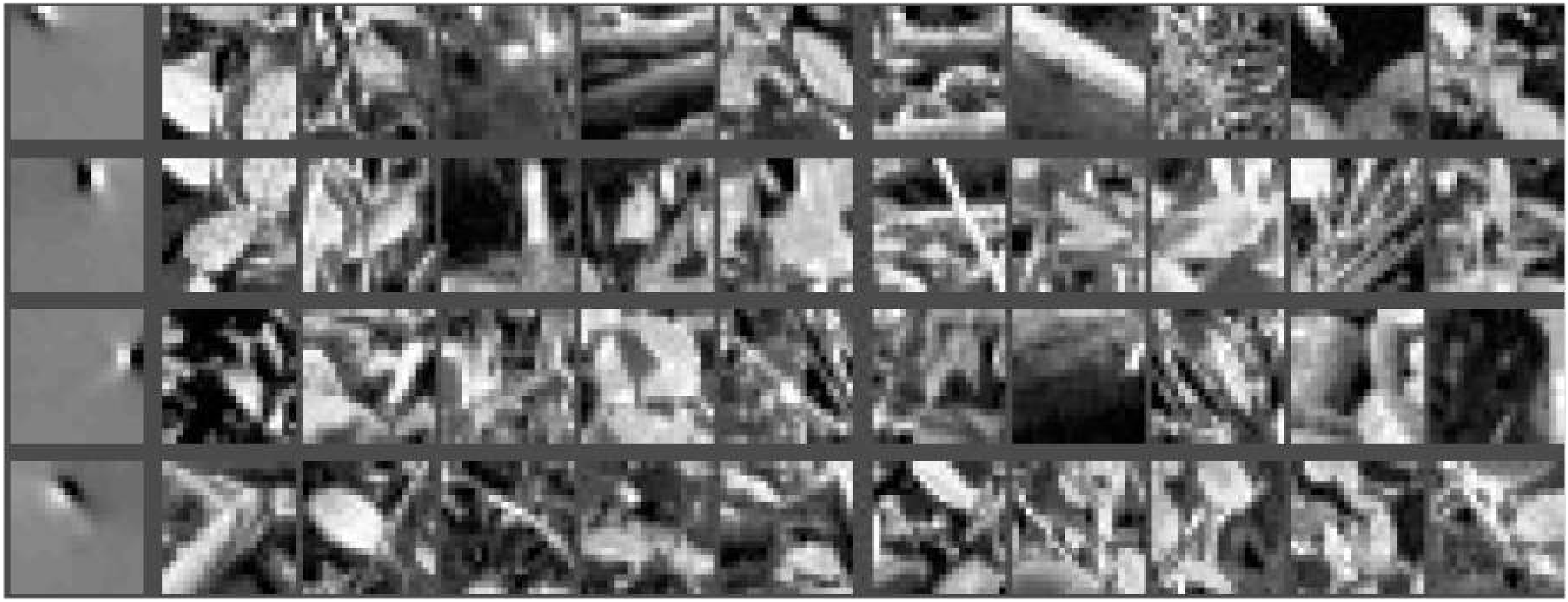
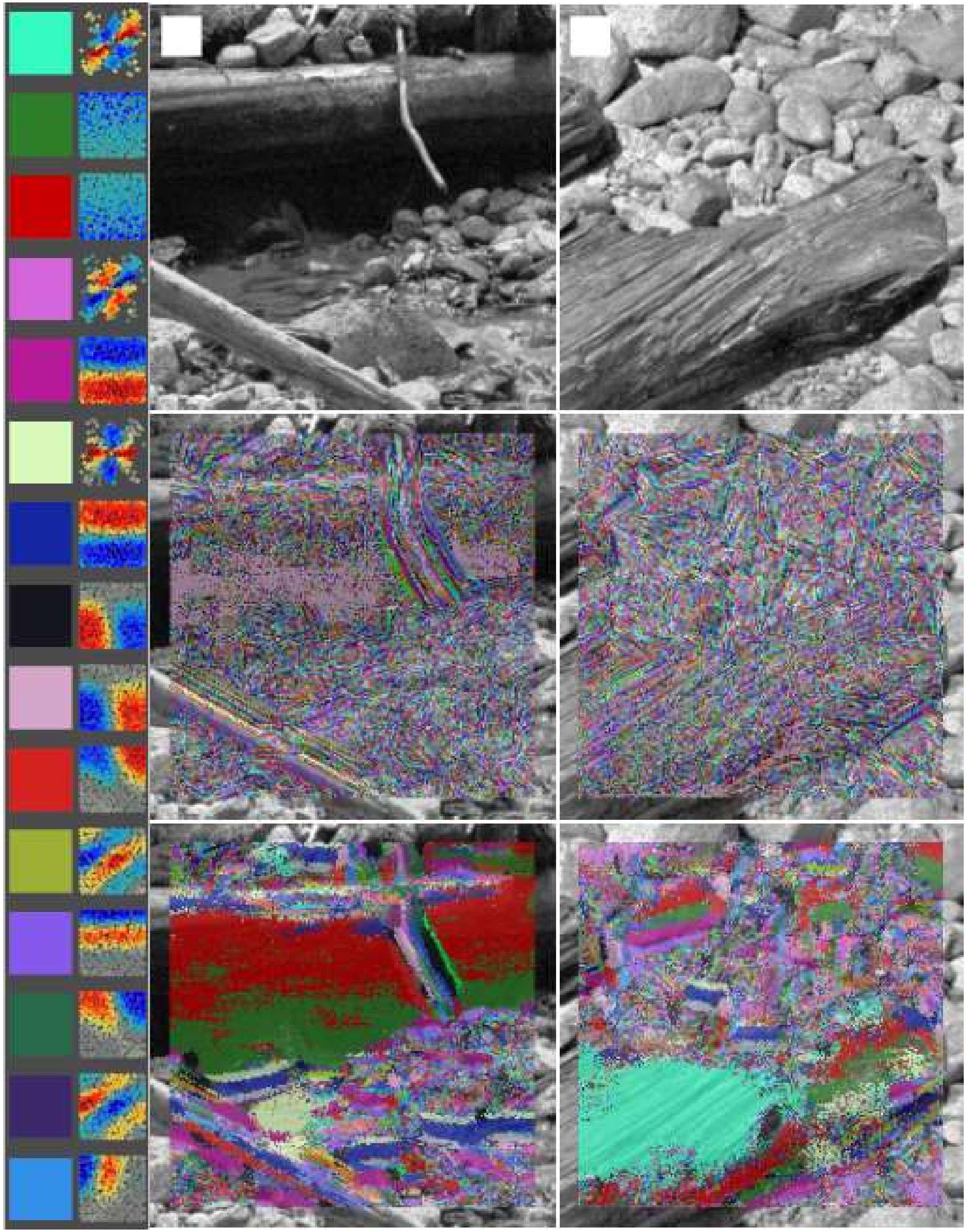


Figure 8



Interpretation

- v_j's are analogous to complex cells, pooling output of first-order features
- model learned “a coarse code of position using broadly tuned spatial patterns”

Comparison to other methods

- Multi-layer sparse coding network (H&H)
  - different nonlinearity
- Gaussian scale mixtures (W&S)
  - trees? local neighborhoods?