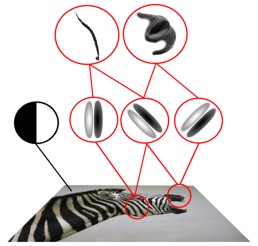
Sampling in a hierarchical model of images reproduces top-down effects in visual perception

Mihály Bányai, Gergő Orbán

Computational Systems Neuroscience Lab, Wigner Research Centre for Physics, Budapest, Hungary

The Component Scale Mixture model of images

- The visual system is representing a hierarchical **generative model** of the environment.
- V1 simple cell responses are organised by **latent variables** representing higher-order statistics of sensory input.
- The latent structure determining covariance structure of V1 cells corresponds to **Gestalt principles**.
- Full **Bayesian inference** is assumed in the model, posteriors are represented by stochastic **samples.**



$$p(v \mid g) = \mathcal{N}(v; 0, \sum_{j=1}^{K} g_j C_j)$$

$$p(x \mid v, z) = \mathcal{N}(x; zAv, \sigma_x I)$$

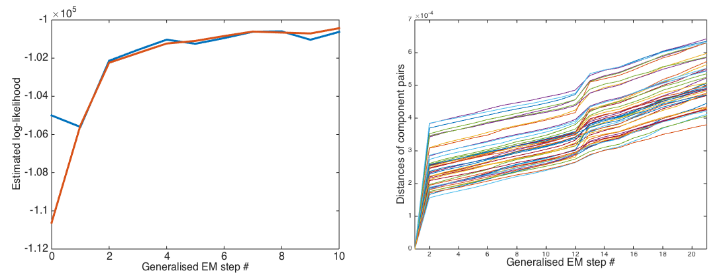
Learning the components

- Generalised EM scheme with gradient ascent
- Averaging over posterior samples in the E-step

$$C_v = \sum_{k=1}^{K} g_k U_k^T U_k \qquad [U_k]_{i,j}^{new} = [U_k]_{i,j}^{old} + \epsilon \frac{\partial \mathcal{L}}{\partial [U_k]_{i,j}}$$

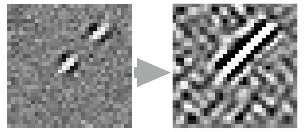
$$\frac{\partial \mathcal{L}}{\partial [U_k]_{i,j}} = \sum_{l=1}^{NL} \mathrm{Tr}\left[\frac{\partial \log p(x^l, v^l, g^l \mid U_{1\ldots K})}{\partial C_v^l} \frac{\partial C_v^l}{\partial [U_k]_{i,j}} \right]$$

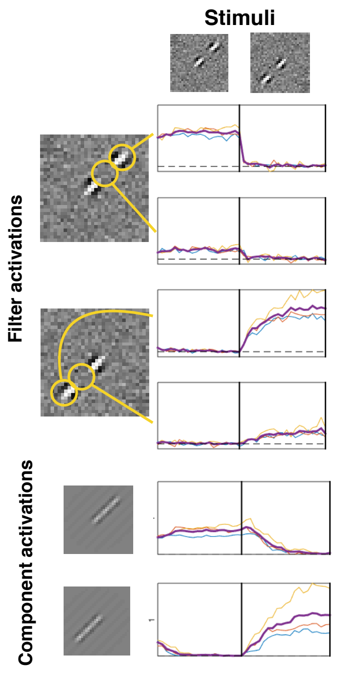
- Log-likelihood of a restricted set of natural images increases with EM-steps
- Each step separates the components from each other

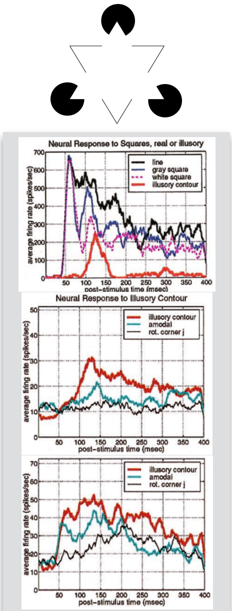


Predicted response to illusory contours

- IC responses are elicited by top-down effects of covariance component activations
- Temporal ordering of activation in latent layers are reproduced by sampling the posterior
- Measured firing rate ratios are reproduced by a synthetic model



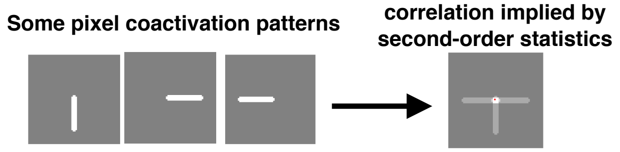




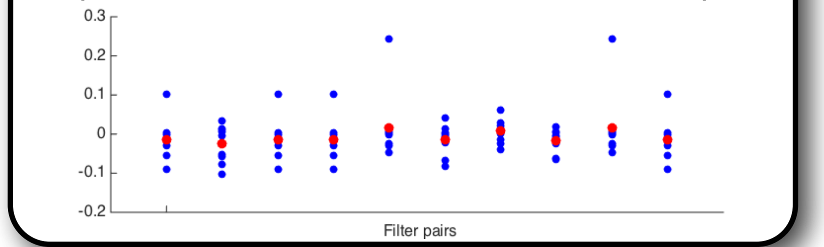
from Lee & Nguyen PNAS, 2001

Relation to GSM

- Wainwright & Simoncelli, NIPS, 2000
- GSM describes the second-order statistics of coefficients
- Selective activation of correlation patterns is only possible with latent variables switching between them
- Learnability of context-dependent correlations also relies on a latent layer preventing the correlations to average out


Some pixel coactivation patterns
correlation implied by second-order statistics

Dispersion of learned correlations around the mean over all components



Sampling the posterior

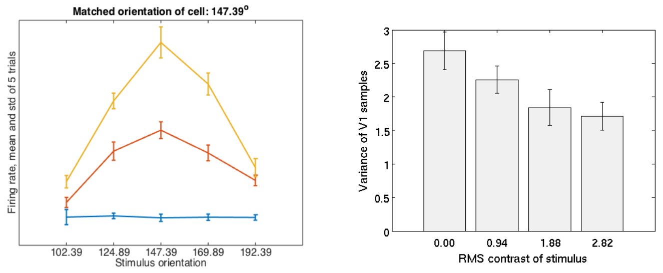
- Generalised Gibbs sampling over the conditional posteriors
- Samples are used to predict membrane potential of cells

$$p(v \mid x, g, z) = \mathcal{N}(v; \frac{z}{\sigma_x} C_{cp} A^T x, C_{cp}),\ C_{cp} = \left[\frac{z^2}{\sigma_x} A^T A + \left[\sum_{j=1}^{K} g_j C_j \right]^{-1} \right]^{-1}$$

$$\log p(g \mid x, v, z) \sim -\frac{1}{2} \left[\log \left(\det \left(\sum_{k=1}^{K} g_j C_j \right) \right) + v^T \left(\sum_{k=1}^{K} g_j C_j \right)^{-1} v \right] + \log p(g)$$

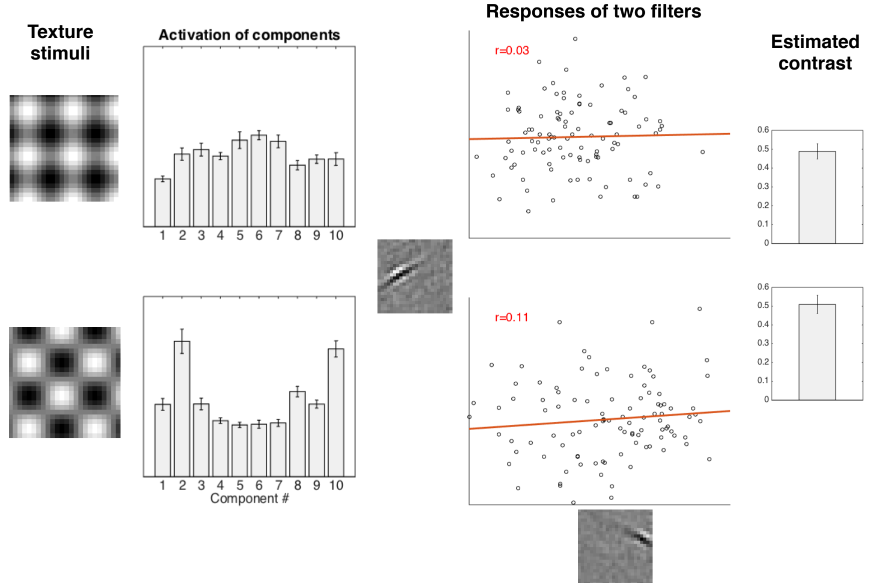
$$\log p(z \mid x, v, g) \sim -\frac{1}{2} \left[D_x \log(\sigma_x) + \frac{1}{\sigma_x} (x - zAv)^T (x - zAv) \right] + \log p(z)$$

- Orientation tuning is independent of stimulus contrast
- Variance of samples decreases with stimulus contrast



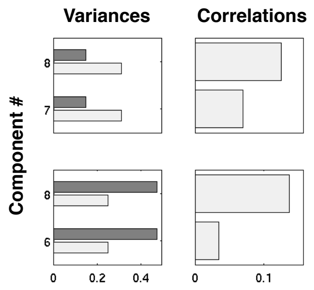
Correlations implied by natural statistics

- CSM model with 10 components using filters from the Olshausen-Field model
- Trained on 24x24 whitened patches from the Vanhateren image database

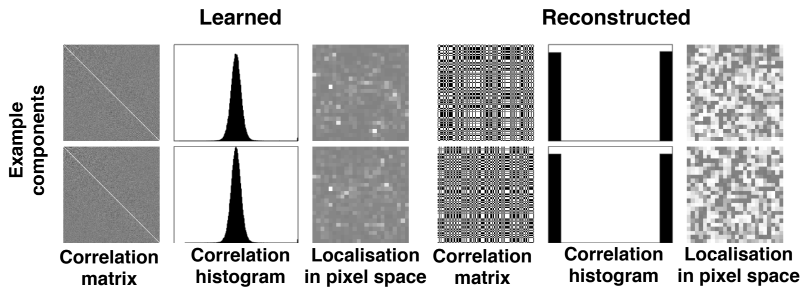


Relation to component models

- Karklin & Lewicki, Nature, 2008, similar model structure
- In CSM we do full posterior sampling instead of giving a point estimate of the latent values, and explicitly represent contrast similarly to GSM
- Parametrisation allows the independent learning of variances and correlations in components



- If we reconstruct correlations from learned variances assuming K&L type parametrisation, we obtain different correlations than through learning them explicitly



Conclusions

- Contextual effects on perception are formalised in a generative model of images
- Sampling from the full posterior enables predictions about variance and covariance
- The model gives predictions for noise correlations between V1 simple cells when fitted to natural image statistics
- The model predicts V1 responses to illusory contours
- CSM generalises GSM and previous component-based image models

Acknowledgement

This work was supported by an MTA Lendület Fellowship