Simple and complex cells as style and content variables in a bilinear model based on temporal stability

Pietro Berkes, Richard Turner, and Maneesh Sahani - Gatsby Computational Neuroscience Unit

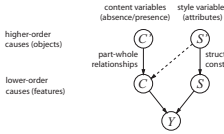
Introduction

Representation of the environment in the sensory cortex:
- How is it structured?
- Which principles underlie its organization?

Basic assumption:
- The goal of the sensory system is to reconstruct the external causes of the sensory input, which is in the end the information needed to guide behavior => **The internal representation should mirror the basic semantics and structure of the environment**.
- Its organization should thus be consistent with some prior knowledge about the basic properties of the external causes

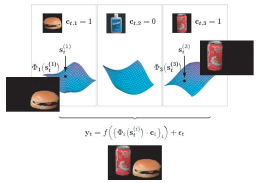
Previously proposed prior structure:
- Independence, sparseness (Olshausen & Field, 1996; Bell & Sejnowski, 1997; Hyvärinen & Hoyer, 2000)
- Temporal stability, predictability (Rao & Ballard, 1999; Hurri & Hyvärinen, 2003; Körding *et al.*, 2004; Berkes & Wiskott, 2005)
- Spatio-temporal "bubbles" (Hyvärinen *et al.*, 2003)
- Bilinear sparse model with global shift variables (Grimes & Rao, 2005)

We propose a model based on:
- **Discreteness and persistence in time** of objects
- Duality of **identity** (absence/presence of an object or feature) and **attributes** (position, orientation, viewpoint, ...). These two aspects have different semantics and should be modeled accordingly. Both are necessary to build invariant representations and to bind attributes that refer to a single object together.



- Conceptually similar to the what/where segregation of the visual stream
- This representation might also be related to some psychophysical effects, like for example temporal versions of the tilt illusion (see proof-of-concept in the bottom right corner)

Bilinear model



Here we define Φ_i to be linear functions, and f to be the sum of its arguments, which leads to the bilinear mapping:

$$\Phi_i(\mathbf{s}^{(i)}) = \mathbf{W}_i\,\mathbf{s}^{(i)}$$
$$\mathbf{y} = \sum_i \mathbf{W}_i\,\mathbf{s}^{(i)}\,c_i + \epsilon$$

(Tenenbaum & Freeman, 2000, Grimes & Rao, 2005, Dayan, 2006)

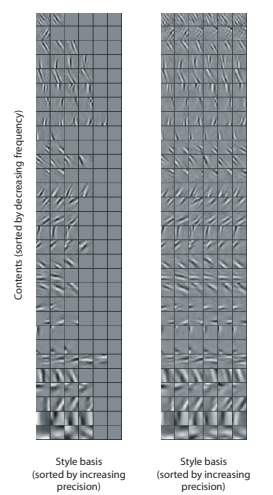
Contents are assumed to be independent but individually persistent, with styles that vary smoothly over time.

We formulate this model in a **probabilistic framework**, which allows us:
- to keep the number of assumptions to a minimum
- to represent uncertainty about the inferred status of the environment
- to learn the size of the model, i.e. number of contents and dimensionality of the content manifold

Simulation

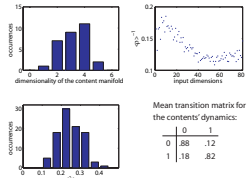
Input data: subset of the **CatCam videos** (Betsch et al., 2004) - several minutes of recordings taken from a camera mounted on the head of a freely-behaving cat. Observations consist of the pixel intensities in fixed windows of size 20x20 pixels.

Simulation results



96 directions remain (slightly overcomplete representation)

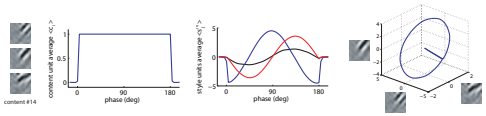
Statistics of the posterior distribution over parameters



Mean transition matrix for the contents' dynamics:

	0	1
0	.88	.12
1	.18	.82

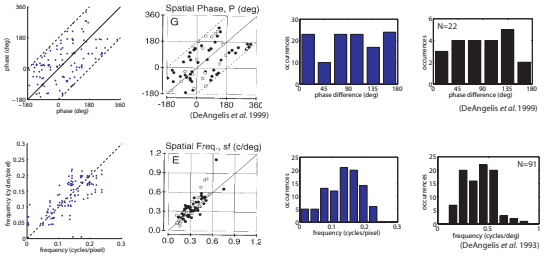
Response to drifting sine gratings

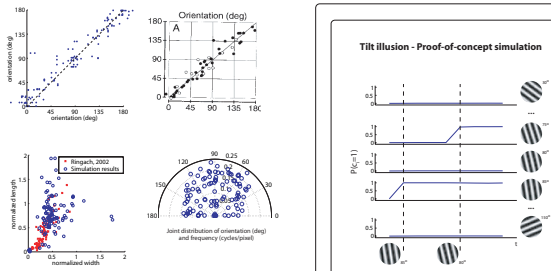


Content and style variables can be interpreted as complex and simple cells in V1. The model suggests that these form two parallel cell populations (as opposed to the classical hierarchical view) and have two distinct functional roles.

Receptive field statistics

In the plots below, we identify the mean of the style variables with the firing rate of simple cells. Simulation data is reported for style variables in the same subspace, while the physiological data from DeAngelis *et al.* (1999) is for simple cells on the same electrode.

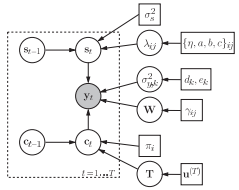




Tilt illusion - Proof-of-concept simulation

Model details

Instead of learning the Maximum Likelihood solution for the parameters, we adopt a Bayesian analysis and integrate (approximately) over a distribution of model parameters. This can be combined with an **Automatic Relevance Determination prior** over the weights in order to learn the size of the model, i.e., the number of contents and the dimensionality of the style manifolds (Beal, 2003).



Contents are modeled as independent, binary Markov chains:

Content model

$$P(C) = \prod_i \left(P(c_{1,i}) \prod_{t>1} P(c_{t,i}|c_{t-1,i}) \right)$$
$$P(c_{t,i} = a | c_{t-1,i} = b) = T_{b,a}\,, \quad a, b \in \{0, 1\}$$

Styles are modeled as Linear Dynamical System, with mean and variance parameters coupled such that the prior variance is 1:

Style model

$$P(S) = \prod_i \left(P(\mathbf{s}_1^{(i)}) \prod_{t>1} P(\mathbf{s}_t^{(i)} | \mathbf{s}_{t-1}^{(i)}) \right)$$
$$P(\mathbf{s}_t^{(i)} | \mathbf{s}_{t-1}^{(i)}) = \mathcal{N}_{\mathbf{s}_t^{(i)}}(\mathbf{\Lambda}\mathbf{s}_{t-1}^{(i)}, \mathbf{\Sigma}_s)$$

With dynamical parameters

$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{d_s}), \quad \text{with} \quad \lambda_1 > \ldots > \lambda_{d_s} > 0$$
$$\mathbf{\Sigma}_s = \mathrm{diag}(1 - \lambda_1^2, \ldots, 1 - \lambda_{d_s}^2)$$

(Turner & Sahani, 2007)

The prior on the generative weights is Gaussian with mean zero:

$$P(\mathbf{W}) = \prod_{ab} P(\mathbf{W}_{:ab}) = \prod_{ab} \mathcal{N}\left(\mathbf{0}, \mathrm{diag}\left(\gamma_{ab}\right)^{-1}\right)$$

After an initial phase, we start learning the precision parameters. For style dimensions that are redundant or not useful, the precision diverges to infinity, forcing the corresponding style basis vector to 0. This then provides an automatic determination of the dimensionality of each content manifold.

We choose conjugate priors for the rest of the parameters, set to be rather noninformative for the input noise and more informative for the dynamic parameters, favouring persistent contents and slowly-varying styles.

Learning is performed using Variational Bayesian Expectation Maximization (VBEM) (Beal, 2003): Inference is performed keeping the whole distribution over parameters and latent variables (as opposed for example to zero-temperature EM). The posterior joint distribution is made tractable by factorizing it into tractable factors, the key VBEM factorization being the one between model parameters and latent variables:

$$Q(\mathbf{W}, \Sigma_y, \lambda, \mathbf{T}, S, C) = \prod_{ab} Q(\mathbf{W}_{:ab}) Q(\Sigma_y) Q(\lambda) Q(\mathbf{T}) \cdot \prod_{t=1}^{T} \prod_{a=1}^{d_s} Q(c_{t,a}, \mathbf{s}_t^{(a)})$$

We introduced three additional factorizations: between the distribution of weights belonging to different contents, between weights and input noise, and between different contents at different times. All other factorizations arise naturally.

References

M. Beal, *Variational Algorithms for Approximate Bayesian Inference*. PhD. Thesis, Gatcby Computational Neuroscience Unit, University College London, 2003.

A. J. Bell and T. J. Sejnowski, *The 'independent components' of natural scenes are edge filters*. Vision Research 37 (23), 3327-3338, 1997.

P. Berkes and L. Wiskott. *Slow feature analysis yields a rich repertoire of complex cell properties*. J Vis, 5(6):579–602, 2005.

B.Y. Betsch, W. Einhäuser, K.P. Körding, and P. König. *The world from a cat's perspective*. Biol Cybern, 90:41-50, 2004.

P. Dayan. *Images, Frames, and Connectionist Hierarchies*. Neural Comp 18: 2293-2319, 2006.

G.C. DeAngelis, G.M. Ghose, I. Ohzawa, and R.D. Freeman. *Functional micro-organization of primary visual cortex: Receptive field analysis of nearby neurons*. J Neurosci 19(9):4046-4064, 1999.

G.C. DeAngelis, I. Ohzawa, and R.D. Freeman. *Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development*. J Neurophys 49(4):1091-1117, 1993.

D.B. Grimes and R.P.N. Rao. *Bilinear sparse coding for invariant vision*. Neural Comp, 17(1):47–73, 2005.

J. Hurri and A. Hyvärinen. *Simple-cell-like receptive fields maximize temporal coherence in natural video*. Neural Comp 15(3), 663-691, 2003.

A. Hyvärinen and P. Hoyer. *Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces*. Neural Comp 12(7):1705-1720, 2000.

A. Hyvärinen, J. Hurri, and J. Väyrynen. *Bubbles: a unifying framework for low-level statistical properties of natural image sequences*. J Opt Soc A, 20(7):1237-1252, 2003.

K.P. Körding, C. Kayser, W. Einhäuser, and P. König. *How are complex cell properties adapted to the statistics of natural scenes?* J Neurophysiol, 91(1):206–212, 2004.

B.A. Olshausen and D.J. Field, *Emergence of simple-cell receptive field properties by learning a sparse code for natural images*. Nature, 381:607-609, 1996.

R.P.N. Rao and D.H. Ballard, *Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects*. Nat Neuro 2, 79-87, 1999.

J.B. Tenenbaum and W.T. Freeman. *Separating style and content with bilinear models*. Neural Comp, 12(6):1247–1283, 2000.

R. Turner and M. Sahani. *A maximum likelihood algorithm for SFA*. Neural Comp, (in press), 2007.