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, UCL

A promising approach to the study of representation in the visual system is based on the idea that neuronal response properties reflect an internal model of the statistical structure of the environment, and can thus be derived or learnt from the structure and parameters of a statistical model of sensory observations, following the Helmholtzian notions of perception as an inferential process. The generative modelling framework is particularly appealing in this respect, because it defines the internal model by making explicit how the visual image is generated from external sources. Successful learning of the models parameters thus implies the identification of such causes, which are ultimately what is needed to control behaviour.

Here, we present a model of visual input where observations are generated from a set of binary ”content” variables, each representing the presence of a particular visual feature. The appearance of each feature is modelled by a multidimensional manifold that represents its episodic pose, and is parametrised by a set of ”style” variables. Contents are assumed to be independent of one another but individually persistent, with styles that vary smoothly over time. Content and style variables then combine bilinearly to form the image. The temporal priors on the latent variables allow us to overcome the degeneracies present in the bilinear model and to perform learning in a fully unsupervised way.

When this model is applied to natural image sequences, the resulting content manifolds are spanned by pairs of localised Gabor wavelets with similar position, orientation, and frequency, but with a 90 degree phase relationship. Consequently, the inferred activity of style variables in the learnt model resembles that of simple cells in the primary visual cortex. Content variables indicate the presence of a feature irrespective of its position on the underlying manifold, and are therefore insensitive to the phase of the corresponding wavelets. Their behaviour is thus similar to that of complex cells.

Thus, this generative model makes explicit an interpretation of complex and simple cells as elements in the segmentation of a visual scene into independent features along with a parametrisation of their episodic appearance. It also indicates their possible role in a hierarchical system that extracts progressively higher-level contents, starting from simpler, low-level features.

We are currently investigating the possibility that the segmentation in persistent contents underlies visual effects such as the tilt illusion: Due to the assumptions of the model, small changes in the input are explained away by the currently active contents while other content variables have to account for the residuals, which leads to a form of repulsion similar to that seen in temporal versions of the tilt illusion.

Acknowledgements
This work was supported by the Gatsby Charitable Foundation.