

# Learning Invariant and Variant Components of Time Varying Natural Images Using a Sparse, Multiplicative Model

Charles F. Cadieu, Bruno A. Olshausen

Redwood Center for Theoretical Neuroscience, University of California, Berkeley

A remarkable property of biological visual systems is their ability to infer structure within the visual world. In order to infer structure, a useful representation should separate the invariant from the variant information [1, 2]. Invariant information is important for determining ‘what’ we are seeing, recognizing objects and interpreting scenes; while variant information captures the ‘where’ or ‘how’ information, the transformations of objects. It has been hypothesized that biological visual systems represent ‘what’ and ‘where’ visual information in two separate cortical processing streams [3]. How do biological systems decompose visual information into separate invariant and variant representations?

To explore such a decomposition, we present a model that learns to separate the invariant from the variant part of time varying natural movies. We first reformulate the sparse coding model [4], in which images are represented as a generative model of linear over-complete bases with sparse causal variables, so that images are instead represented in terms of a multiplicative interaction between two sets of causal variables. One set of variables is constrained to change slowly over time (the invariant representation), and the other set of variables is allowed to change quickly over time and is encoded as a phase angle (the variant representation). These variables effectively decompose the original sparse coding variables into invariant and variant representations.

After training on natural image sequences, the learned basis functions are similar to those produced by the original sparse coding model: Gabor-like functions that are spatially localized, oriented and bandpass. However, the multiplicative decomposition produces both invariant components with slowly changing responses, which indicate the presence of a visual shape, and variant components in the form of processing phase angles over time, which indicate their transformations. The model thus predicts two classes of cells in primary visual cortex that form the beginnings of the ‘what’ and ‘where’ cortical streams. Moreover, our model’s decomposition provides a starting point for the construction of hierarchical models that capture the global structure and interaction of the ‘what’ and ‘where’ representations.

## Acknowledgments

This work was supported by NGA grant MCA 015894-UCB and NSF grant IIS-06-25223.

## References

- [1] Separating style and content with bilinear models. J. B. Tenenbaum, and W. T. Freeman, *Neural Computation* 12(6), 1247-1283, 2000.
- [2] Bilinear sparse coding for invariant vision. D. Grimes and R. Rao, *Neural Computation* 17(1), 47-73, 2005.
- [3] Two cortical visual systems. L. Ungerleider, and M. Mishkin, In D. Ingle, M. Goodale, and R. Mansfield (Eds.), *Analysis of visual behavior*, 549-585, (1982). Cambridge, MA: MIT Press.
- [4] Emergence of simple-cell receptive field properties by learning a sparse code for natural images. B.A. Olshausen and D.J. Field, *Nature* 381: 607-609, June 1996.