Discovering Compositional Structure

by
Matthew T. Harrison
B.A., University of Virginia, 1998
Sc.M., Brown University, 2000

Doctoral dissertation
Ph.D. Advisor: Stuart Geman

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy
in the Division of Applied Mathematics at Brown University

Providence, Rhode Island
May 2005

Chapter 5

Future work

5.1 Learning selectivity and invariance

The experiments in Chapters 3 and 4 illustrate that selectivity can result from incorporating spatial dependencies into a probabilistic graphical model and that invariance can result from incorporating temporal dependencies. In future work we hope to address the original goal set forth in Chapter 1 of combining both into a hierarchal framework, specifically, a hierarchy of reusable parts, or a composition system.

There are several major hurdles. The first and foremost is probably computation, although many recent examples in the literature have demonstrated the feasibility of computing with large graphical models using a variety of different techniques. (For some specific examples, see [5, 7, 6].) We do not address computation here. A more theoretical concern is how to learn selectivity on top of invariance. Of course the solution to this problem undoubtedly affects computation.

5.2 The Markov dilemma

Invariance, by definition, hides information. Translation invariance hides location. Rotation invariance hides orientation. Scale invariance hides size. In the context of statistical computing, invariance can be viewed as a Markov assumption. The decisions made on top of a translation invariant representation are conditionally independent of actual location information. Location information is accumulated (and thereby lost) by the invariant layer before this information is passed on to further layers. Markov assumptions form the backbone of all probabilistic graphical models and it is not clear how to proceed with computation and estimation in their absence.

Furthermore, as we mentioned in Chapter 1, the right type of invariance seems crucial for fast learning and generalization because it can substantially lower the dimensionality of the state space. Of course, it has to be the right type of invariance. If the information that needs to be learned is hidden by an earlier level of invariance, then learning – fast or slow – is impossible. And therein lies the problem: how do we know what it is the right type of invariance? Stuart Geman and Elie Bienenstock refer to this as the *Markov dilemma*: Markov assumptions (i.e., invariance) seem necessary for computation and learning, yet

Markov assumptions often hide information that is needed for later decisions. In particular, invariance seems likely to hide information that is needed for later selectivity. Note that this affects both computation and learning. Note also that it is peculiar to hierarchies of invariance and selectivity. Figure 5.1 is their canonical illustration of the problem [1]. Trying to build a long (translation invariant) bar detector out of two adjacent, smaller translation invariant bar detectors does not necessarily work. The large bar detector cannot tell when the smaller bars are appropriately aligned.

The Markov dilemma is closely related to the well known *binding problem* in neural systems [8, 9]. (See the special issue of *Neuron*, Vol. 24, Sept. 1999, for a collection of discussions about the binding problem.) Presumably such a problem will present itself when we try to combine the two approaches from Chapters 3 and 4. Some of the avenues that we will likely need to explore are mentioned below. They are certainly not mutually exclusive.

- **No dilemma.** If we think about selectivity as arising from spatial dependencies and invariance as arising from temporal dependencies, it is not clear why there should necessarily be a conflict. A better theoretical understanding of this connection might reveal a good solution. For example, the ideas of sparsity and entropy reduction that have played such an important role in methods like sparse coding and ICA, which are essentially learning selectivity, can also be applied to the temporal stability methods for learning invariance. Indeed, a binary random sequence is sparse (i.e., has low entropy) not just when 1's are rare, but also when 1's clump together (temporal stability). Perhaps an information theoretic framework that unified the two would also give insight into the Markov dilemma.

- **Over-representation.** A common solution when a useful Markov assumption is too strong is to increase the state space. Consider, for example, the situation in Figure 5.1 where the information in the two smaller invariant detectors is not sufficient to distinguish one long bar from two short bars. There are differences between the two cases other than location, however. The case with two bars also has features like end-of-bar or parallel-bars. Similarly the case with one bar also has a variety of short bars in between (spatially) the two short bar detectors. If all of these types of features were also represented by the system, even if they were represented invariantly, it might be possible to build an invariant long bar detector not with just two short bars, but perhaps with many short bars and also with the absence of end-of-bars and parallel-bars. By building a long bar detector out of many more features (increasing the state space) we might be able to overcome the information that is lost by the invariance. (See [4] for a discussion and examples of this strategy for object detection and/or recognition. See also [6] for an example where location information is explicitly included in the state space.)

  The drawback, of course, is that increasing the state space typically makes both computation and learning more difficult. It is not clear how large the state space will need to be in order to create a recognition system that is robust to clutter. Another drawback is that suspicious coincidence detection is overwhelmed in a densely distributed representation (see Chapter 2 and the references therein), so our model building strategy would likely have to be modified.

- **Auxiliary information.** It might be possible to partially violate the Markov property and still preserve much of the benefits. For example, certain summary auxiliary information like location and scale could be passed around the invariance. The difference between this idea and increasing the state space is that presumably this auxiliary information would be treated fundamentally differently, perhaps augmenting a more classical approach based on the Markov property.

5.3 Neural systems

Another approach is to investigate how the brain might surmount the Markov dilemma, or to see if it even needs to. An obvious first step is to gather more detailed information about the response properties of visual neurons. Many groups have and continue to do just this. Nevertheless, the technical and statistical hurdles are tremendous. It is still not clear if the caricature from Chapter 1 is an accurate picture. In Chapter 6 we suggest some statistical techniques for investigating the response properties of visual neurons in a more agnostic manner. Agnostic methods are becoming more and more popular for making precise statements about the amount and type of selectivity and invariance in the visual system (see Chapter 6 for references).

Working on the assumption that the visual hierarchy exists, there have been several proposed solutions to the Markov dilemma (or variants of it, like perceptual binding) that loosely fit into our "auxiliary information" category above, for example, the use of sophisticated attention mechanisms [3]. Several of the proposed solutions in the literature are based on the fine temporal structure of neural firing patterns. (See [2] for a recent review of the best known variants of this idea.)

In the specific context here, Stuart Geman has proposed a solution that uses partially synchronous firing to pass information about functional connectivity through the invariance [1]. This information is sufficient for distinguishing the two cases in Figure 5.1. In Chapter 7 we describe some jitter-based statistical techniques for investigating the temporal resolution of neural firing patterns in a more agnostic manner. Jitter methods offer a great deal of flexibility for investigating things like synchronous firing, but with few modeling assumptions.

Bibliography

[1] Stuart Geman. Invariance and selectivity in the ventral visual pathway. 2004 (in preparation).

[2] Charles M. Gray. The temporal correlation hypothesis of visual feature integration: Still alive and well. *Neuron*, 24:31–47, September 1999.

[3] John H. Reynolds and Robert Desimone. The role of neural mechanisms of attention in solving the binding problem. *Neuron*, 24:19–29, September 1999.

[4] Maximilian Riesenhuber and Tomaso Poggio. Are cortical models really bound by the "binding problem". *Neuron*, 24:87–93, September 1999.

[5] Stefan Roth and Michael J. Black. Field of experts: A framework for learning image priors with applications. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005 (submitted).

[6] A.J. Storkey and C.K.I. Williams. Image modelling with position-encoding dynamic trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):859–871, 2003.

[7] Jian Sun, Nan-Ning Zheng, and Heung-Yeung Shum. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):787–800, 2003.

[8] C. von der Malsburg. The correlation theory of brain function. Internal Report 81–2, MPI Biophysical Chemisty, 1981. Reprinted in *Models of Neural Networks II*. E. Domany, et.al., eds. Springer, Berlin, 1994.

[9] C. von der Malsburg. Binding in models of perception and brain function. *Current Opinion in Neurobiology*, 5:520–526, 1995.

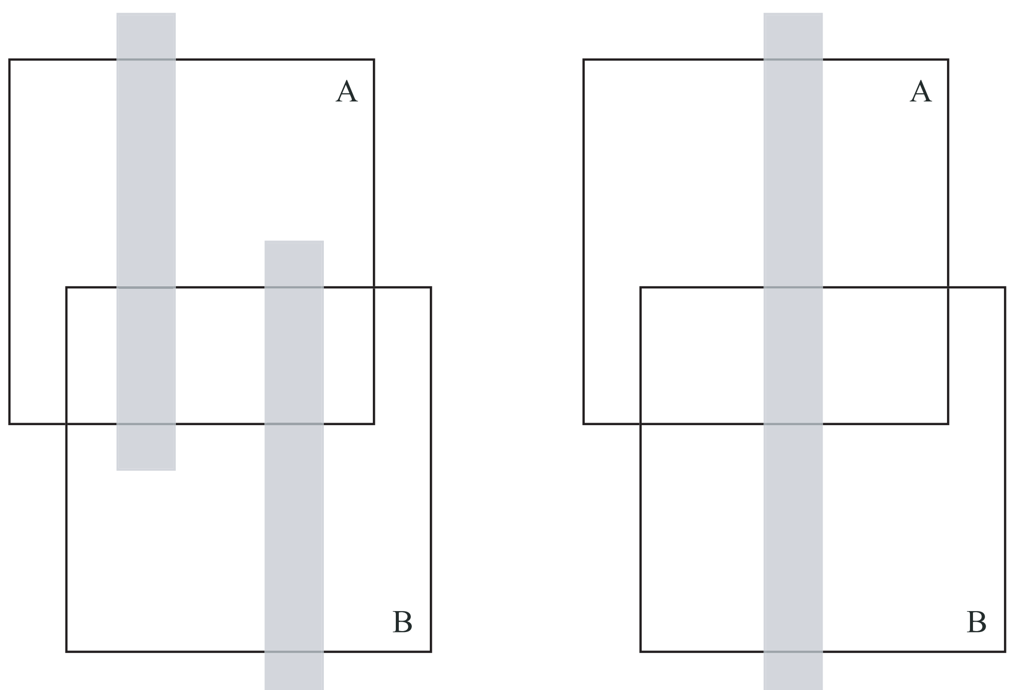

Figure 5.1: This figure from [1] illustrates the Markov dilemma. An invariant bar detector responds equally to any vertical bar in box A. Another responds equally to any vertical bar in box B. The situations on the left and the right cannot be distinguished based on the responses of these two filters alone. If we wanted to use these two invariant bar detectors to build a longer invariant bar detector we would need extra information. The invariance of the small detectors hides information that we need for later selectivity.