# The Modal Workshop

Commentary on 'The perception of shading and reflectance,' by E. Adelson and A. Pentland

Allan D. Jepson
University of Toronto

March 14, 1995

Adelson and Pentland[1] discuss the issue of selecting a particular interpretation for a given image from the set of interpretations which are consistent with the image data. Indeed, the authors argue that there are often several alternative interpretations for any single image and that it is the perceptual system's job to choose one or more preferred interpretations from within this set of possibilities. To motivate their approach to this selection problem the authors present an elegant metaphor of a workshop whose task involves minimizing the overall cost of reconstructing the scene. The interpretation that minimizes this cost over the set of consistent interpretations is chosen as the 'percept' for the image. This same issue of selecting a low cost or preferred interpretation from the set of possible interpretations is addressed in each of the other chapters in this book, although the currencies used in the other chapters are not workshop dollars. Rather the other chapters consider cost functions based on probability estimates or, in some cases, concepts built on top of probability, such as utility and risk.

In view of the fact that this selection problem appears throughout this book, it is useful to reformulate Adelson's and Pentland's workshop metaphor in terms of the common Bayesian framework used elsewhere. The specific problem domain addressed by the target chapter, namely that of interpreting shape, pigmentation, and lighting from images, provides an excellent concrete example of many of the issues brought out in other chapters. Moreover, the reformulation of the workshop raises a few points about Bayesian models which have not been highlighted so far. My intention with the reformulation presented below is to keep the general spirit of their metaphor, as developed in their fee schedule, but not necessarily the details of the reported algorithm.

[1] This commentary is to appear in the book "Perception as Bayesian Inference" edited by D. Knill and W. Richards. See Sinha & Adelson, ICCV'93 for material related to the target chapter.

# 1 Modal Specialists

The most direct way to construct a Bayesian model is to replace the three specialists in the workshop metaphor with stochastic specialists. For example, the stochastic painter would, with a particular probability, spontaneously paint a rectangular patch. Also, with a lower probability, it can paint a general polygon. Moreover, by setting the cost of an operation (according to the reported fee structure, say) to be proportional to the negative logarithm of the probability of that operation occurring in our stochastic workshop, we can directly associate costs of producing an item in the target chapter's workshop with the probability of the item being randomly produced by our stochastic workshop. That is, in the stochastic workshop the probability of independently selecting operations to form a sequence is just the product of the probabilities of selecting each operation. By identifying the cost to be proportional to the (negative) log probability we see that the total cost of the sequence is then just the sum of the costs for each step. Note that we use the negative of the log probability so that high costs are associated with scenes that are produced only rarely by the stochastic workshop.

These first steps in mapping the workshop to a stochastic model immediately bring forward one aspect of Bayesian models which has so far been neglected in this book. For example, our percept of the zig-zag shape depicted in the target chapter includes a description of the artifact's structure and a description of some of the steps in a process for generating that structure. The important point, vividly brought out by the workshop metaphor, is that our percepts not only include a description of what is in the scene and where things are, but also at least a partial description of a process for the construction and placement of the various objects in the scene (see [1]).

Returning to our reformulation, consider the structure of the probability distributions appropriate for each stochastic specialist. In particular, we argue that it is natural to take these probabilities to have a modal structure, as defined in Chapter 4. For example, there is a significant probability that a rectangle is painted, rather than a more general four sided polygon. Moreover, the sides of a painted rectangle can be made parallel to other lines within the artifact. Similarly, cuts are straight, often at right angles to other cuts, and bends can be precisely at right angles. In all these cases we have events occuring with a nonzero probability, but which exist on smaller dimensional sets than some more general embedding space. Thus our stochastic workshop is indeed a 'modal workshop' in that the prior distributions for the basic operations performed in the workshop have a modal structure. Therefore an approach of the form considered in Chapter 4 could be used to specify the modal structure of the various specialists.

The presence of modal properties suggests that various reliable inferences can be made from image data. The flip side of this coin is that some sets of interpretations should be extremely improbable. A classic example of an improbable inference is shown in Figure 11.5 of the target chapter. Given the processes assumed to exist in our modal workshop, we might expect an extremely small probability for generating the depicted arrangement of three strips, arranged in a precise way, and viewed from a particular position so that the image matches that in Figure 11.2. For example, given that there is no mode for arranging the leading edges of each of the three strips to be coplanar, and so on, these structures would have to arise simply by chance. This is improbable and, since a small probability corresponds to a large cost, we expect the cost of the structure depicted in Figure 11.5 to be extreme.

Interestingly, this extreme cost is not reflected in the given fee schedule. Notice there is no cost for the precise positioning of various parts, nor for positioning the observer, nor for making accurately specified cuts or folds. A consequence of the lack of charging for accidental views, and so on, is that there can be some surprising minimal cost solutions within the provided fee schedule. A simple example is provided by an image consisting of a rectangle partitioned into a dark square and a light square. The minimal cost solution is the spatial expert's solution ($15 for a rectangle bent at right angles, illuminated with a flood light and viewed from an accidental position). Our percept is instead of a painted rectangle ($18, by painting only one of the squares, along with using a flood light and a fronto-parallel view). The lack of charges for making improbable inferences given the assumed modal structure of the domain therefore appears to be an oversight in the fee structure suggested in the chapter. The alternative approach of starting with a probabilistic model, for which the various modes are included in the basic formulation, appears to be a principled way to keep track of such large costs.

# 2 Hidden Costs

Some of the large costs associated with modal structure do appear implicitly in the algorithm suggested for solving the problem, even though they are not explicitly listed in the fee structure. For example, the difficulty with the interpretation of the half-painted rectangle, as discussed in the previous paragraph, is avoided in the proposed algorithm through the use of the constraint that lines which are straight in the image are actually straight in the scene. This is done by allowing the surface model depicted in Figure 11.6 to have corner beads only at places which are not junctions of two colinear line segments. By setting up the problem in this way the algorithm effectively implements a hard constraint that colinear line segments in the image are necessarily colinear in the scene.

Alternatively, one might imagine that the algorithm imposes an arbitrarily large cost on breaking this colinearity constraint, which is related to an extremely low probability within our stochastic workshop. But does a Bayesian model support such a property? In fact, the appropriate Bayesian analysis of the colinearity inference is given by the discussion of key features presented in Chapter 4. This analysis shows that, given an appropriate modal prior distribution, colinear line segments in the image provide reliable evidence for the colinearity of the corresponding lines in the scene. That is, the probability in favour of the colinear interpretation can indeed be extreme or, equivalently, the costs associated with breaking such a constraint can be taken to be extreme. Therefore it may seem natural to rigidly impose this constraint in the actual formulation of an algorithm.

But the same Bayesian analysis of key features shows that the situation is not quite that simple. In particular, it highlights the requirement that there must be an appropriate modal prior, and this raises serious concerns about using a built-in colinearity constraint. In fact,

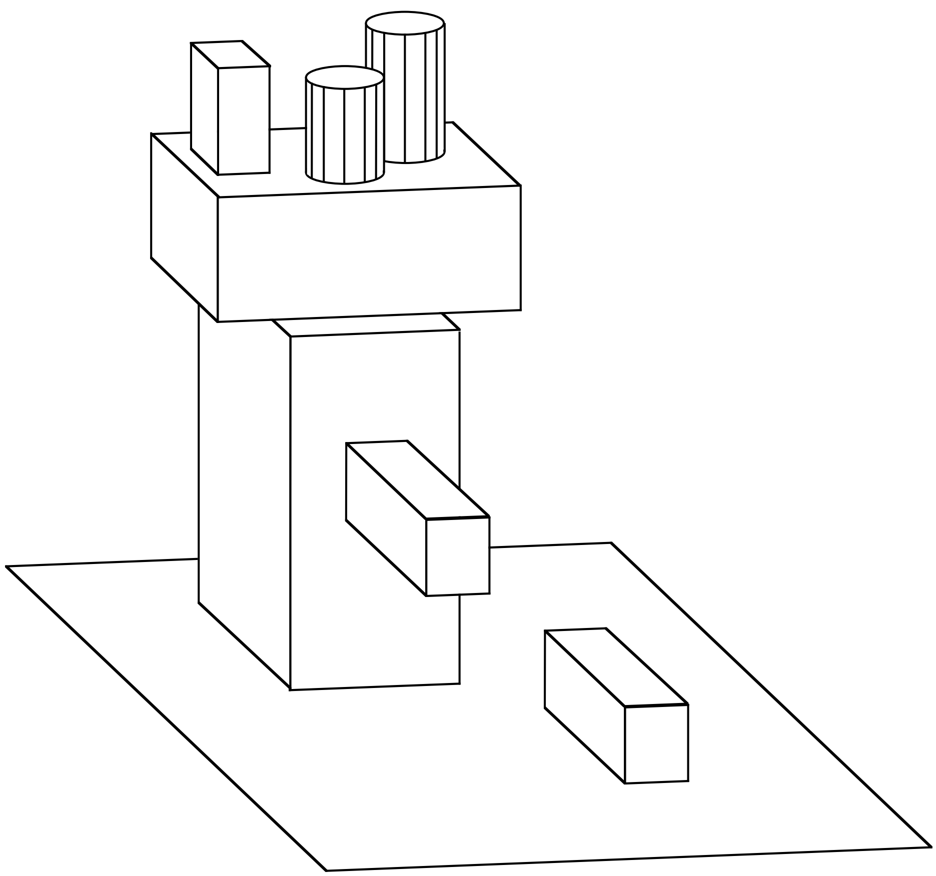

Figure 2.1: An accidental view of an accidental scene. In this image there are four pairs of colinear line segments, plus a set of three colinear segments, none of which are perceived to be colinear in the scene.

there are two problems with built-in constraints such as "colinear lines in the image are colinear in the scene", both of which are illustrated in Figure A. The first problem is that they represent an over-commitment to what is essentially a probabilistic inference. Given such a hard constraint, no amount of contradictory evidence can succeed in causing it to be retracted. An example of a situation in which contradictory evidence appears to override the colinearity property is given by the depiction of the two cylinders in Figure A. Note that even though the two cylinders are shown to share a line segment, our perceptual systems can break the colinearity constraint in favour of the evidence for two separate cylinders resting on a flat surface.

The second difficulty with using a built-in colinearity constraint is that such a hard constraint ignores the possibility that the required modal prior may be conditional on other aspects of the interpretation. Again this situation occurs in the preferred interpretation of Figure A. In particular, note that there are five different colinear alignments between edges of different blocks, none of which are perceived to be colinear in the scene. In each case, the missing modal property would require a block to be floating freely in space but nevertheless be precisely positioned to have one of its edges colinearly aligned with the edge of another block. Such a mode can be safely assumed not to exist. Moreover, without this mode the appropriate inference to make is that the line segments should be interpreted as accidental alignments of non-colinear segments in the world (see Chapter 4). That is, the odds are strongly against the colinear pair of lines in the image actually being colinear in the world, but rather favour the viewer being accidentally aligned. (The detailed Bayesian analysis for this case is similar to the discussion in Chapter 4, and is left to the reader.) Thus our common perception of Figure A shows that, at least in some situations, our perceptual systems can effectively take the intricacies of the appropriate Bayesian analysis in stride.

# 3 The Stochastic Supervisor

In addition to the specialists, Adelson and Pentland also introduce a supervisor. In their metaphor the supervisor simply charges a flat rate whenever there is cooperation amoungst the specialists. For the modal workshop we imagine a stochastic supervisor which randomly chooses a sequence of operations (according to some constraints on what operations are feasible for the various specialists). The supervisor then issues this sequence to the specialists. Given these commands, the stochastic specialists would then perform the operations according to probability distributions which are conditioned by the commands. For example, returning to our stochastic painter, when it is told to paint a rectangular patch it could choose the location, orientation, size, and aspect ratio from some probability distribution, and then paint the chosen rectangle to within some random error. Alternatively, the supervisor could completely specify the rectangle, and the only randomness introduced by the painter would be errors in precisely executing the command. In either case we see that the appropriate charge to associate with the supervisor alone is just the negative log probability of it generating a particular sequence of commands. Therefore the flat rate charged by the supervisor in the target chapter can be interpreted as a stochastic supervisor which simply randomly selects a command sequence from a large finite set of equiprobable sequences.

The stochastic supervisor and modal specialists together define the prior distribution for the sequences which the workshop can execute. This distribution, in turn, specifies the prior distribution for the various artifacts that are produced by the workshop. But how should we view these distributions? Clearly, with the supervisor simply picking a sequence of operations from a huge set of equiprobable possibilities, the model does not describe a competent prop department of any particular theatre company. Presumably the prop department would have been asked to make something like a set of steps before coming out with the zig zag shape shown in the target chapter. Do we need to incorporate such higher level constraints into our stochastic model of the prop department's workshop? If one was attempting to model the annual output of the prop department of a particular theatre company then the unavoidable answer is that we do need to consider the sorts of props required, the sorts of plays performed, and the various styles of staging used. In other words, if the probabilistic model is meant to describe the world then all these factors appear relevant. However, as pointed out in Chapter 1, there is a second way to view these prior distributions. In particular, the modal workshop described above can be viewed as a specification of a perceiver's model of what to expect in its world. That is, the distributions are not meant to accurately model the world, but rather serve to specify the perceiver's model for what is more or less probable in its world (see Chapter 1). If we pursue this second line of reasoning the remaining question is, of course, how does the perceiver get away with using the wrong priors? Why doesn't the perceiver exhibit significant biases or even hallucinate objects according to its incorrect prior distribution?

There are three inter-related reasons that a perceiver might be able to function appropriately with incorrect priors. The first is that, because of the structure in our world and the information available about that structure from a typical image, there may be a large amount of evidence available for a given interpretation and this evidence dominates any bias introduced by the prior model (see Chapter 4 and Chapter 8). The second reason is that the priors used by the perceptual system can be expected to be considerably less structured than priors that more accurately describe a particular domain in the world. Here the stochastic supervisor provides a perfect example, with any one of a huge set of processes treated as equally probable (see also Chapter 13 where the priors are taken to be uniform). In contrast, we might expect the supervisor in a theatre company's prop department to make some choices much more often than others. By choosing a flatter prior the perceiver may avoid the introduction of strong biases simply by avoiding the introduction of unwarranted structure within its prior distributions. The cost of using flatter priors is that, in cases where more detailed and structured priors are available, the perceptual system will not be as statistically efficient as it might be (see Chapter 10 for a discussion of efficiency). Finally, the third reason a perceptual system may perform adequately given incorrect priors is that, in cases where there are several different solutions with comparable posterior probabilities (or costs), the perceptual system need not commit to the single most probable (or minimal cost) interpretation but rather might choose to explicitly represent ambiguities through the appropriate choice of resolution for various parameters (see the use of loss functions for light source location in Chapter 8). Moreover, further ambiguity may be explicitly represented by attempting to provide all the categorically distinct interpretations having roughly comparable posterior probabilities (see Chapter 6 for an approach which emphasizes this).

## 4 Can your supervisor add?

This latter point about uncertain priors raises the possibility that the analogue of Adelson's and Pentland's supervisor, which attempts to minimize the overall cost, simply cannot add. The issue has been discussed in Chapter 6, where we considered the several modes for the construction and placement of a handle. In order to compare the probabilities of interpretations incorporating one or another mode, it was shown to be necessary to know the relative probabilities of the priors for these modes, along with measurement resolutions, and so on (see Table 6.4a). One approach to this problem would be to pick a flat prior, such as is used by the stochastic supervisor discussed above. A second approach is to treat the actual prior as uncertain, but nevertheless attempt to obtain an ordering of the various possible interpretations despite this uncertainty. For an example of this second approach, assume that the three specialists in Adelson's and Pentland's metaphor use three different currencies, one for each specialist, and for which there are no known exchange rates. In such a case the supervisor is faced with three bottom lines, with no information about how to convert to a common currency.

Indeed the algorithm actually implemented by the authors is perhaps more appropriately thought of in terms of precisely these three distinct currencies. Recall that in the implemented algorithm the shape expert gets the first crack at minimizing costs, then the lighting expert has a chance, and finally the painter touches things up. As a result, the effective cost function being minimized by this algorithm treats any savings created by the shape expert as more valuable than any extra expenses that might be incurred by the lighting and paint specialists. The supervisor is therefore minimizing the costs reported by the three experts by using a priority ordering, treating the spatial expert as first priority and the painter last. We note in passing that this nicely illustrates how the specification of an algorithm might radically change the effective cost function which is being used to compare various interpretations.

The ability to compute a single 'bottom line' through the use of a common currency, or a priority scheme, is central to the workshop metaphor presented in the target chapter and also to the Bayesian approaches described in most other chapters of this text. Such a cost function, after all, forms the basis of the ordering of the set of possible interpretations. Having access to such a cost function may appear to be both simple and intuitive, at least from the perspective of a theoretician attempting to model perception. But is it simple from the perceptual system's point of view? To get a glimmer of what this might look like to our perceptual system, imagine that we had just such a cost function to make common decisions. For example, who should I hire for that research assistantship? Why can't we just crunch some numbers and get a simple total ordering of the candidates? The knowledge requirements that would go into computing such a cost function are, of course, overwhelming. It is for just this reason that in Chapter 6 we attempt an alternative approach to the problem of ordering the various possible interpretations for the purpose of selecting one or more as the 'percepts' of a given image.

# References

[1] M. Leyton. A process grammar for shape. *Artificial Intelligence*, 34(2):213–247, 1988.