From Features to Perceptual Categories

W. Richards*

J. Feldman A. Jepson

First we review an analysis of conditions that should be met if features are to provide robust inferences about world properties. Features meeting these conditions provide indices into especially useful categories of visual properties. Then we show that for a given set of elemental concepts the categories associated with these properties have a natural hierarchical (specialization) structure. We argue that this structure provides constraints on the form and type of categories that are inferred when visual objects are classified.

1 Introduction

Perception, or "seeing", involves the assignment of world properties to image elements. Both machine and biological vision systems proceed in this task by recasting the image pixels into meaningful "features", from which object properties are inferred [1]. The features suitable for this task are not arbitrary, but are highly constrained, and have a natural hierarchical structure. This structure mirrors that of natural processes, thereby providing a basis for inferring natural categories. Hence we begin with a review of conditions that constrain the choice of those special features that provide robust inferences about world properties.

2 What Makes a Good Feature?

Let the world consist of various properties P that are associated with various contexts, C. Then $p(P|C)$ denotes the conditional probability of a property, P, such as "has 4 corners" in the context C, which could be sitting "on a plane", "in this region", etc. Similarly the collection of measurements of a property and their conditional probabilities will be specified by F and $p(F|C)$. Note that $p(P|C)$ and $p(F|C)$ are simply objective facts about the world and are *not* statements about the perceiver's model of the world. Our first task is to place conditions on F, P and C that ensure the measurements F constitute a reliable indicator that P occurs in the world.

***Authors contributed equally and their order is arbitrarily permuted. WR and JF are at Mass. Inst. Tech., Dept. Brain & Cog. Sci., Cambridge, MA 02139; AJ is at the Univ. Toronto, Dept. Computer Science, Toronto M5S 1A4.** **This work was supported by AFOSR 89-504 and NSERC Canada.** **Correspondence should be addressed to WR.**

BMVC 1992 doi:10.5244/C.6.11

2.1 Reliable Inferences

The posterior probability of inferring property P given the feature F in context C is $p(P|F\&C)$. A reliable inference makes this probability nearly one, and keeps the probability of an "error", i.e. $p(notP|F\&C)$ near zero. Hence a reliable feature F, in context C, will keep the following ratio, namely R_{post} much larger than one:

$$R_{post} = p(P|F\&C) \ / \ p(notP|F\&C) \,. \tag{1}$$

Using Bayes Rule, R_{post} can be broken down into the product of two components, a likelihood ratio L that relates to the "imaging" of P onto F, and the prior probability R_{prior}, that relates to the genericity of the world property P in context C. Specifically, $R_{post} = L \cdot R_{prior}$, where

$$R_{prior} = p(P|C) \ / \ p(notP|C) \ \text{ and } \ L = p(F|P\&C) \ / \ p(F|notP\&C). \tag{2}$$

Note that the likelihood ratio captures the intuition that a feature should arise reliably from a given world property, i.e. $L >> 1$. As will be seen in the next section, however, this condition does not insure a reliable inference, because if R_{prior} becomes too small, then R_{post} can become insignificant even in the presence of a high likelihood ratio. (Also see [2].)

2.2 An Example

Consider a world of line segments on a plane seen under orthographic view. Of interest is the special property "two line segments are parallel". Let the threshold for discriminating the orientation difference between two (adjacent lines) be θ, and let $\delta << \theta$ be the limiting resolution of the process that governs straight and parallel. Now let the collective distribution of the orientation ϕ of all line segments be rather flat (Figure 1A). Given this context, we are now presented with two lines that fall within the crosshatched sample for $\phi < \theta$. Hence the two lines appear parallel; should we conclude that these lines indeed arise from a parallel process?

First note that the likelihood ratio, L, is very high, because (i) whenever parallel lines occur in the world, they always will appear parallel in the image, and (ii) our chance of error is vanishingly small – when two lines are not parallel, they will not be seen as such except in the rare case when they lie within our limit of resolution θ. Hence $p(F|P\&C) = 1$ and $p(F|notP\&C)$ is, say 0.01 if θ is 1 part in 100. It appears therefore that we should infer that the lines are indeed parallel in the world. However,

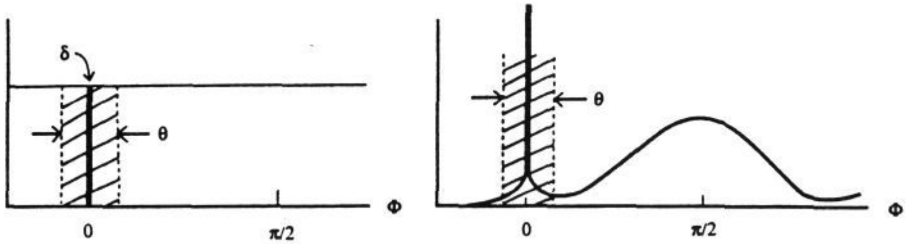

Figure 1 A: Flat distribution function. B: "Modal" distribution.

given our chosen random world context, such an inference is almost always guaranteed to be *wrong.*

Consider the prior probability ratio R_{prior}. Because the prior probability δ of the parallel process occurring is much less than the resolution limit θ, the area occupied by δ in Figure 1A is much less than the area set by θ. Thus $R_{prior} = \delta/(1-\delta) << \theta$, and its product with the likelihood $L \cong 1/\theta$ will give an a posteriori probability ratio $R_{post} < 1$. Hence the odds really favor the conclusion "not parallel". (See [2] and [3] for further details and examples.) In order to raise R_{post} to a significant level, we need significant priors, say a δ in this case such that $\delta/\theta >> 1$. In terms of Figure 1, this is equivalent to requiring that the ϕ distribution function for pairs of lines be biased, such as indicated in Figure 1B where the process "parallel" appears as a mode in the probability distribution function.

3.0 Model Class

The important message of the previous example is that "good" features arise from some modal regularity in the distribution function of world properties. However, not all regularities satisfying the likelihood and prior conditions will be useful. For example, the property "two skewed lines" satisfies these two conditions, but clearly this property is not very informative. Hence what we seek are properties that are not just arbitrary configurations, but rather ones that are in some sense special.

3.1 Two Kinds of Regularities

Structural regularities within a given model class can be divided into two classes: transverse and non-transverse [4]. Transverse relations arise when the elements of the model are postioned arbitrarily such as the above two skewed lines; non-transverse arrangements require careful positioning, as implied by the term "non-accidental" of Binford [5] and Lowe [6]. Unlike the notion of "non-accidental", however, the usage of transverse and non-transverse requires a context. Thus, "two parallel lines" (or planes) in a random stick (or planar) world would be non-transverse, but in the context of a building with windows and doors, etc., the concept "parallel" would become transverse. Within the proper context, non-transverse properties are thus very special. But as we showed earlier, in order to be recoverable from image features, the non-transversality must be an isolated spike in the distribution function as in Figure 1B, with sufficient mass to be "visible". This is what previous researchers meant by "modal" properties [7, 8, 9]. Features that satisfy (2) and which arise from non-transverse regularities provide especially reliable and useful inferences about world properties and are called Key Features (see [2] for "natural" examples taken from motion and color). Loosely speaking, F will be a Key Feature for property P if P is a generic non-transverse mode in the space of world models, and F occurs in the presence of P but never in its absence. Hence the set of properties that image onto the Key Features are an especially useful set of properties, because they are reliably inferable.

3.2 An Example

To illustrate a set of properties that image onto key features in our simplified world of line segments in a plane, assume there are two processes that generate two types of relations between two lines. One is the process "parallel"; the other is a process "coincident", where the lines just touch one another. We take these regularities as generic – i.e. we stipulate that both occur with significantly non-zero probabilities in the given context. First we enumerate those regularities that image to key features. Then in the following section, we will place an ordering on this special set of properties.

The enumeration is equivalent to identifying all the non-transverse configurations between line segments in a plane, given the chosen context. We assume the measurement is the orientation of one line to the other, ϕ, and the position x, y of the end-point of one line with respect to the other. Hence the relative positioning has three degrees of freedom (DOF). Referring to Figure 2, the uninformative, transverse regularity chooses x,y, and ϕ arbitrarily, producing two skewed lines. (Intersection or not was not specified in our model class and an "X" will be treated as equivalent to skew without crossing.) First, with care the end of one line can be placed

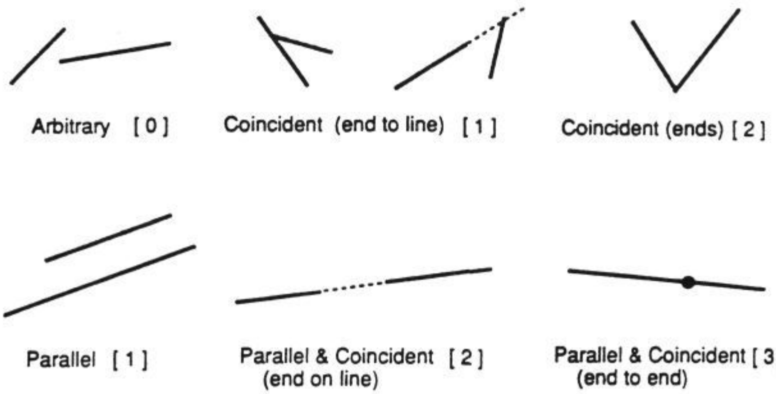

Figure 2 Line-to-line non-transversalities.

on the other (or its extension), eliminating one degree of freedom. These configurations are assigned a codimension of one. Next, with still more care, we can place the end of one line exactly on the end of the other, allowing only the angle ϕ to vary. This arrangement has a codimension of two.

Similarly, if the lines are parallel, then the orientation is fixed and the cost, or codimension of the arrangement is again one. However parallel and coincident lines, with one end allowed to slide along the other, increase the codimension further to two. Finally, we have the last, most special case of positioning of codimension 3 where the two lines merge into one when placed end to end in a parallel arrangement.

4.0 From Features to Categories

Our main point will be that the "interesting" structural regularities in a model class – namely those that satisfy the key feature conditions – can be used as a basis for partitioning the model class into categories. In the previous example, the property space would be built from the end-point position measurements x, y, and the relative pose ϕ. Within this x, y, ϕ space, our proposal is that the partioning should make explicit the line-to-line non-transversalities illustrated in Figure 2. If this scheme is adopted, then the subspaces will preserve the character of the nontransversal modes, thus distinguishing among the interesting properties. Note that the context sensitivity is critical to our set-up, because it permits legitimate reconfigurations of the property space depending upon the observer's goals, etc.

4.1 "Two Stick" Categories

Continuing our example, we identify the modal subspace as that associated with our "two-stick" model class presented earlier in Figure 2. Each of these configurations has a codimension, which allows us to place each of these non-transverse modes in a lattice, where each node depicts a proper subspace in the particular context (Figure 3A). The top node shows the arbitrary two-stick configuration. As we move down the lattice, the nodes below differ at each successive level by the removal of exactly one degree-of-freedom from the configuration. Upward transitions, then, are the elemental ones that locally "break" or "unfold" a non-transversal property but which do not add any additional non-transverse properties. An important example is the missing link between the "V" and "parallel line" nodes (or similarly, the "T" and "collinear" nodes). There is no direct route from one node to the other. The explanation is that the concepts "coincident" impose a constraint on the endpoint position x, y of one line with respect to the other, whereas the concept "parallel" is expressed by an angular relation, ϕ between the two lines. Because position (x, y) is not defined by angle (ϕ) or vice versa in this context, there is no intersection other than the excluded degenerate case of two coincident lines. A similar explanation applies to the missing path from the two "collinear" lines and the "T" node.

At the bottom of the lattice, two nodes have the two sticks collapsed to one. These two nodes have broken outlines to indicate that they are not part of the lattice for the perceptual context because they suggest a "one-stick" configuration. (If the two sticks were each identified in some manner, say by coloring, then the dashed paths and nodes would become part of this "two-stick" category lattice.)

4.2 "Natural Example": Beetle Lattice

In the biological realm, growth processes exhibit regularities [10]. To illustrate how such regularities can be used for a taxonomic classification, we will use a simple modification of the "two-stick" mode lattice of Figure 3A. Let the context be the backs of beetle-like bugs that are marked by two distinctive lines oriented with respect to the symmetry axis of the beetle. As is typical for biological shapes, we assume the markings are generated symmetrically about this axis. Hence, with respect to our "two-stick" mode lattice, one stick – the "reference stick" – will simply be the symmetrical bisector of the beetle's back. The other, namely the "second stick", will thus

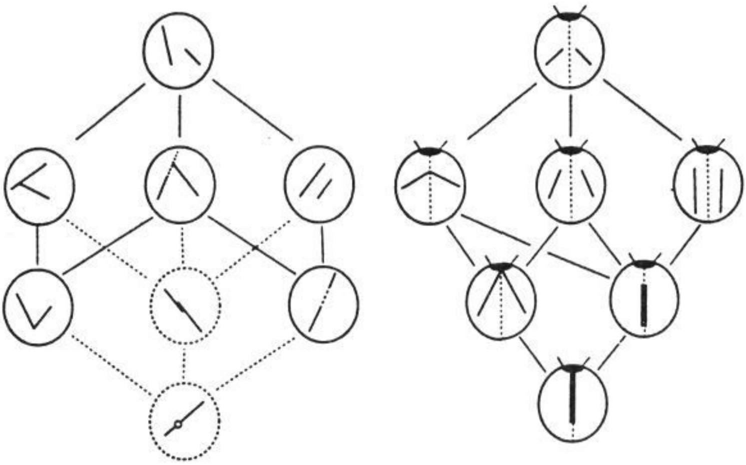

Figure 3 A: "Dot-on-line" categories. "Two-stick" categories given concepts coincidental andf parallel. See text for explanation of dashed paths and nodes. B: Beetle taxonomy, based on a version of a "two-stick" mode lattice.

appear twice in mirror image positions about this symmetric axis. (The situation is equivalent to symmetric markings appearing on a left and right wing.) As before, we assume two possible marking processes, one laying the mark down parallel to the bisecting axis, the other positioning the end point (of either the reference stick axis or the additional marking stick) to be coincident with one of the two lines. All of this sets the context.

Because the two-stick modes in a similar context have already been enumerated, we simply need to recast the previous lattice of Figure 3A in a symmetric form compatible with this revised "biological" context. This has been done in Figure 3B, where now each node depicts the markings on the beetle's back. At the top, the two symmetric marking lines are set arbitrarily with respect to the bisecting axis (dotted). This is the codimension 0 case for this species. At the next level either the coincident or parallel process applies, giving us three codimension 1 subspecies. Next, we have two codimension 2 cases: in one the marking lines form a V, coming together at the "head" of the reference line, or the other where the two marking lines collapse onto the reference line (but do not reach the head of the beetle). Finally we have a single codimension 3 case in this context where the "V" collapses onto the reference bisecting line. Given these generating processes and this context, these are all the types of beetles expected. These types, with the exception of the "generic" beetle at the top, represent the beetle modes or subspecies, each exhibiting a slightly different, but related regularization of the ontology of beetles. Thus the beetle lattice is a convenient hypothesis generator for an observer who is seeking to assign any particular beetle to its "natural" category [11, 12].

5.0 Category Induction

It is easy to imagine that pairs of sticks or beetle markings in a world might obey the regularities depicted in the nodes of the lattice of Figure 3, having all originated from some common underlying processes. Then it would seem plausible that object categories corresponding to the regularities found on the lattice are more prone to occur in the world than completely arbitrary collections of stick-pairs that are not on the lattice (This is the Natural Mode assumption [7, 9, 13].) Putatively, then, when an observer examines a collection of stick-pairs, or beetles, the conclusions about what category processes are responsible for the observed collection will be drawn exclusively from the conceived lattice. Consider the induction problem in Figure 4A. What is the common description of the collection on the left as opposed to the collection on the right? An infinity of different answers are possible, many making entirely different predictions about what "more of the same" would mean on each side. However, the simple answer "V's versus parallel lines" is most intuitively compelling. Note that such predictions about what new examples of either side will look like constitute an enormous inductive leap, since so many alternative solutions are also possible given the small number of exemplars. The problem in Figure 4B can similarly be solved almost instantly, namely T's and V's versus "parallel and collinear lines". Note that our conclusion was not T's versus parallel lines, although the V's and collinear lines are respectively degenerate cases of T's and parallel. Thus each node in the lattice is taken as a separate category in its own right.

The previous example shows that the mixture of two nodes is never on the lattice. This property allows categories to be inferred correctly by an observer even when they are intermingled, because the collapsed compositional category that all objects are of the same kind is not a sanctioned hypothesis (see [14]). Again this is seen in panel C, where now we have mixtures of different beetle subspaces. Hence as long as the world presents

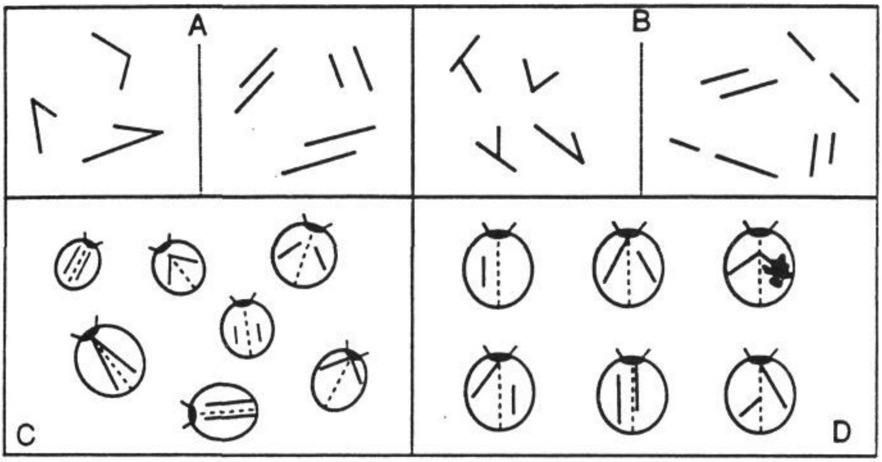

Figure 4 Category induction.

only lattice regularities, an observer can report correctly the number of categories present – a capacity unaccounted for by conventional categorization techniques such as clustering.

Finally, in panel D we have presented a set of beetles that clearly appear defective. Note that these beetles fall outside the normal beetle definition because of their asymmetry. Rather than concluding these forms don't fit, however, instead we attempt to hypothesize that they are altered defective forms of recognizable categories. On the top half of Figure 4D we are successful, in the the bottom half we are not – these latter beetles seem "weird" mutants. Hence whenever possible the lattice is used to "regularize" the defective beetles back to the true form [15]. When such regularization is impossible, then the beetle is seen as inconsistent with our models for the generative process of natural forms. Hence the category lattice built from easily recognizable non-transverse structural modes appears to play a major role in object classification and recognition.

References

[1] D. Marr, Vision: a computational investigation into the human representation and processing of visual information. Freeman, San Francisco, 1982.

[2] A. Jepson and W. Richards, What makes a good feature? To appear in: Harris L and Jenkin M (eds) Spatial vision in humans and robots. Cambridge University Press, 1992. Also available as MIT AI Lab Memo 1356, 1992.

[3] D.C. Knill and D.K. Kersten, Ideal perceptual observers for computation, psychophysics and neural networks. In: Watt R J (ed) Pattern recognition by man and machine. MacMillan London, 1991.

[4] T. Poston and I. Stewart, Catastrophe theory and its applications. Pitman, London, 1981.

[5] T.O. Binford, Inferring surfaces from images. Artif Intell 1981; 17: 205-244.

[6] D. Lowe, Perceptual organization and visual recognition. Klewer, Boston, 1985.

[7] A. Bobick, Natural categorization. MIT Artif Intell Lab Tech Report 1001, 1987.

[8] D. Marr, A theory of cerebral neocortex. Proc R Soc Lond B 1970; 176: 161-234.

[9] W. Richards and A. Bobick, Playing twenty questions with nature. In: Pylyshyn Z (ed) Computational processes in human vision: an interdisciplinary perspective. Ablex Norwood NJ, 1988.

[10] D. Thompson, On growth and form. The University Press, Cambridge, 1952.

[11] J. Feldman, Perceptual simplicity and modes of structural generation. Proc 13th Ann Conf Cog Sci, Chicago, IL, August, 1991, pp 299-304.

[12] M. Leyton, Perceptual organization as nested control. Biol Cybernetics 1984; 51: 141-153.

[13] A. Witkin and J.M. Tenenbaum, On the role of structure in vision. In: Beck J, Hope B and Rosenfeld A (eds) Human and machine vision. Academic New York, 1983.

[14] J. Feldman, Constructing perceptual categories. Proc Comp Vis and Pat Recog, Champaign, IL, June, 1992, pp 244-250.

[15] J. Feldman, A. Jepson and W. Richards, Is perception for real? Proc Conference on Cognition and Representation, SUNY Buffalo, April 1992, in press.