MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL INFORMATION PROCESSING

A.I. Memo No. 1356
C.B.I.P. Paper No. 72

April 1992

What Makes a Good Feature?

W. Richards and A. Jepson

Abstract: Perceptual information processing systems, both biological and non-biological, often consist of very elaborate algorithms designed to extract certain features or events from the input sensory array. Such features in vision range from simple "on-off" units to "hand" or "face" detectors, and are now almost countless, so many having already been discovered or in use with no obvious limit in sight. Here we attempt to place some bounds upon just what features are worth computing. Previously, others have proposed that useful features reflect "non-accidental" or "suspicious" configurations that are especially informative yet typical of the world (such as two parallel lines). Using a Bayesian framework, we show how these intuitions can be made more precise, and in the process show that useful feature-based inferences are highly dependent upon the context in which a feature is observed. For example, an inference supported by a feature at an early stage of processing when the context is relatively open may be nonsense in a more specific context provided by subsequent "higher-level" processing. Therefore, specification for a "good feature" requires a specification of the model class that sets the current context. We propose a general form for the structure of a model class, and use this structure as a basis for enumerating and evaluating appropriate "good features". Our conclusion is that one's cognitive capacities and goals are as important a part of "good features" as are the regularities of the world.

© Massachusetts Institute of Technology 1992

Acknowledgment: This report describes research conducted in the Department of Brain & Cognitive Sciences, the Center for Biological Information Processing, at the Artificial Intelligence Laboratory, and at the University of Toronto Department of Computer Science. WR is supported by AFOSR 89-504 and NSF-IRI8900267, and AJ by NSERC Canada. Support for the Artificial Intelligence Laboratory Research is provided in part by grant S1-801534-2 from Hughes Aircraft Company and by the Advanced Research Projects Agency of the Department of Defense under Army contract DACA76-85-C-0010, and in part by ONR contract N00014-85-K-0124. We appreciate the helpful comments and issues raised by Jacob Feldman, Horace Barlow, Richard Mann and Donald Hoffman.

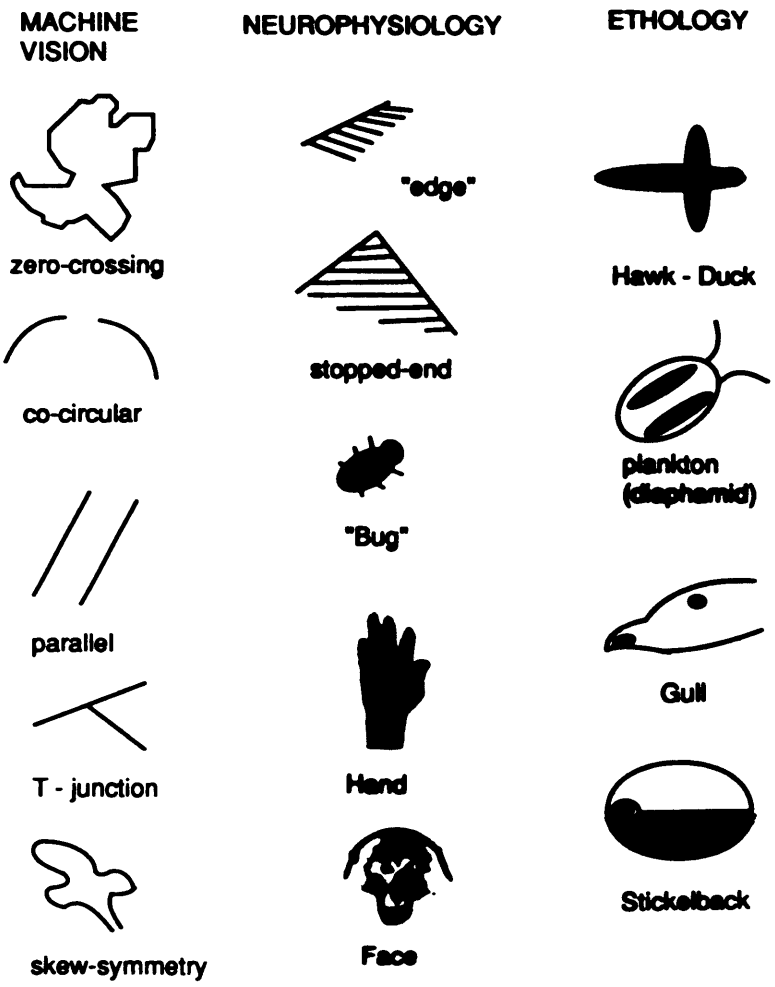


Figure 1 Typical features proposed by machine vision, neurophysiology, and ethology. What common properties do these features satisfy? What makes one feature better than another?

In contrast, consider configurations of features that exhibit very special relations to one another, such as two line segments which intersect to form a "T" or a "V", or two line segments that are collinear. As noted by many (Barlow, 1985; Binford, 1981; Lowe, 1985), intuitively, such coincidences imply very special "suspicious" and informative events. Surprisingly, however, in an unrestricted context, such as a world where sticks are positioned arbitrarily, the observation of a "non-accidental" feature typically does *not* imply the intended world property. Again, context plays a crucial role, as illustrated in Figure 2 for the T-junction, which can arise in many different ways. To correct this situation, the corresponding world event must express a generic regularity in that context (Bennett et

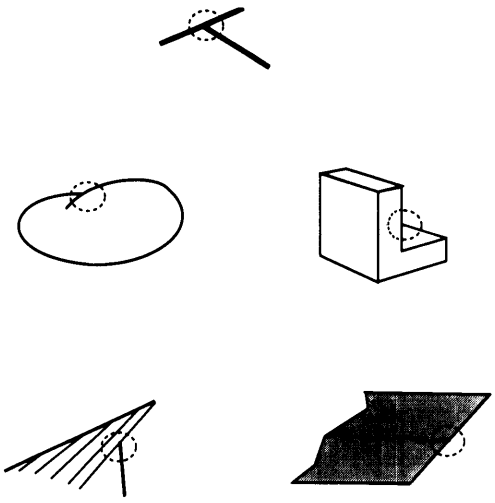


Figure 2 If the image primitives are contours (such as zero crossings), then features typically can be created in many ways. For example, the T-junction may arise either from an occlusion or from an actual T-vertex in 3D. Hence the interpretation associated with a feature depends strongly on the context. Alternate contexts can reverse the interpretation. For example, consider the peanut shape as a wire frame, or the bottom right figure as the view of a crack through a polygonal hole.

al., 1989; Marr, 1970; Reuman & Hoffman, 1986; Witkin & Tennenbaum, 1983). Our task here is to make note of such conditions needed to support our intuitive notions of what 'makes a good feature'. In the process, we will place a measure on just how "good" a particular feature is for inferencing, and show that such measures depend upon the current conceptualization of the world.

2.0 Bayesian Framework

To explore conditions that should be satisfied by a good feature, we use a probabilistic model as the analytical tool for modeling the perceiver's world and the reliability of its feature-based inferences. Our choice of a probabilistic model is *not* a claim that the perceiver necessarily has access to the various probability density functions we use in our analysis. Whether or not the perceiver itself needs to incorporate such a probabilistic model to

distinguish between good and bad features, and whether the world needs to satisfy this particular model, are important issues addressed later in the second part of our proposal regarding the inference process itself. However, a Bayesian probabilistic formalism allows us to state clearly some conditions that a "good feature" should meet, and to explain why other, seemingly obvious proposals are inadequate.

The structure of the model is as follows. The external world consists of different classes of objects and events. We refer to each class as a context, C, within which are various properties that occur probabilistically. Our canonical property is denoted simply by P, and we assume it occurs in context C with the conditional probability $p(P|C)$. We denote the absence of property P by $notP$. Next, we consider that some measurements are taken of the objects and events in the world. We refer to a particular collection of such measurements as a feature F. Hence a feature will be identified with the set of all world events having measurements specified by F, and thus probabilities such as $p(F|C)$ are well defined. We wish to study the inference that property P occurs in the world, given both that the world context is C and that the measurements F are satisfied. Note that the probabilities $p(P|C)$ and $p(F|C)$ are considered to be objective facts about the world (or at least an idealization of the world), and are *not* statements about the perceiver's model of the world. In this section we keep the issue of whether or not a perceiver needs to use any probabilistic model of the world quite separate from our analysis of a good feature.

2.1 Reliable Inferences

In the probabilistic formalism a measure of the success of inferring property P from F is the a posteriori probability of P given the feature F in the context C. A reliable inference makes this probability, namely $p(P|F\&C)$, nearly one, and the probability of an error, namely $p(notP|F\&C)$, nearly zero. It is convenient to consider the ratio of these two quantities, that is

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

We consider the feature F to provide a reliable inference, in the context C, precisely when this probability ratio R_{post} is much larger than one. Below we consider how such a condition can be ensured.

Bayes' rule can be used to break down the probability ratio R_{post} into two components. The first component, L, is a likelihood ratio and relates to the measurement F of property P. The second component is another probability ratio, R_{prior}, and is related to the genericity of the world property P in context C. The decomposition of R_{post} has the simple form:

$$R_{post} = L \cdot R_{prior} \ . \tag{2}$$

Here the prior probability ratio R_{prior} is given by (compare equation (1))

$$R_{prior} = \frac{p(P|C)}{p(notP|C)} \quad . \tag{3}$$

and the likelihood ratio L is defined to be

$$L = \frac{p(F|P\&C)}{p(F|notP\&C)} \quad , \tag{4}$$

From equation (2) we see that the likelihood ratio L acts as an amplification factor on the prior probability ratio R_{prior}. Thus it makes sense that a good feature F have a large amplification factor:

> **Measurement Likelihood Condition**: In context C, a good feature F for world property P provides a large likelihood ratio, that is,
>
> $$L = \frac{p(F|P\&C)}{p(F|notP\&C)} >> 1 \quad . \tag{5}$$

At first blush, a large likelihood value for L seems sufficient to capture the intuition that good features should point reliably to some property in the world. However, because L appears as a product with R_{prior} in equation (2), it is clear that we can not afford to let the prior probability ratio R_{prior} become too small. That is, we also require

> **Genericity Condition**: Given a context C and a constant $\delta > 0$, the property P occurs with probability $p(P|C) > \delta$ or, equivalently,
>
> $$R_{prior} = \frac{p(P|C)}{p(notP|C)} > \frac{\delta}{1-\delta} > 0. \tag{6}$$

By "generic" we mean that P occurs with a probability greater than zero within context C. The Genericity Condition puts a lower bound of δ on this probability. Given that L and R_{prior} satisfy the likelihood and genericity conditions, it follows from equation (2) that $R_{post} > L\delta/(1-\delta)$. Hence, when $L >> (1-\delta)/\delta$, the two conditions together ensure a reliable inference.

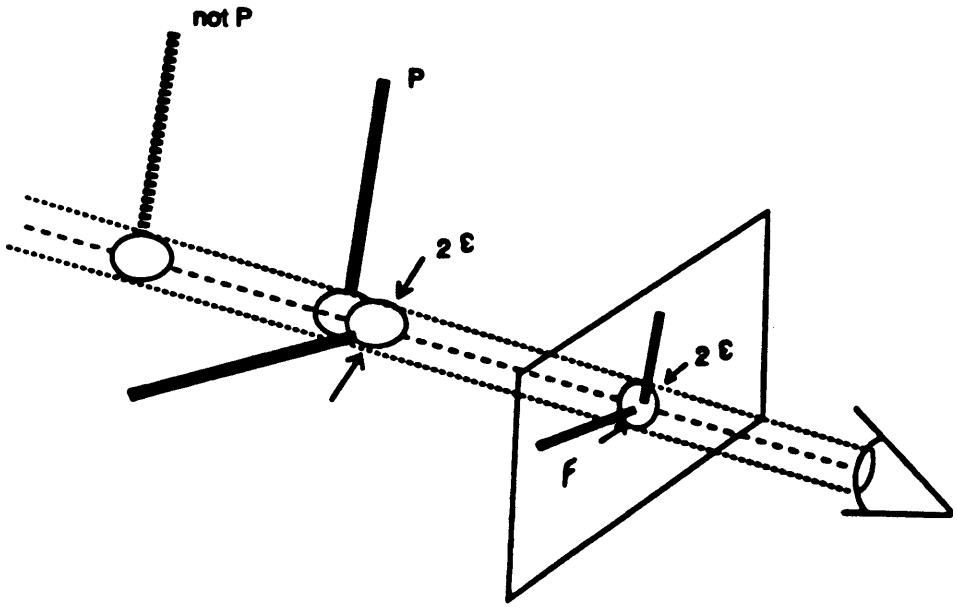


Figure 3 Two sticks in 3D form a near-V vertex to create property P, which projects into the V-junction image feature F. The resolution for the sticks forming a V is taken as a disc of radius ϵ in the image (assuming orthographic projection) and, for the 3D tolerance, the sphere of similar radius. Although the measurement likelihood ratio condition is satisfied, the conditional probability of P, given the observation F and a random world context, favors $notP$ – i.e. that the endpoints of the two sticks lie at separate locations within the cylinder of radius ϵ.

that case $R_{post} = \delta/\epsilon^2 >> 1$. But this is simply the genericity condition, which requires a context in which the 3D "V" structures are fairly common. In other words they are a regularity in that context (Bennett et al., 1989; Marr, 1970; Witkin & Tennenbaum, 1983), such as if we are in a blocks world where edges form V's, or perhaps another where "victory signs" are created by finger arrangements. Once again, then, the context plays a major role in the inferences that features support.

2.3 Informativeness

By requiring that both the genericity condition be satisfied as well as $L >> 1$, we now can be assured that the feature F in context C will be a reliable predictor of world property P. However, a third condition is needed to ensure that the inference of P is actually informative. For example, in a context of randomly placed sticks (e.g. C_{open}) consider a world property P such as two skewed sticks. For simplicity we assume an orthographic image mapping and let the feature F correspond to two skewed lines in the image. Then

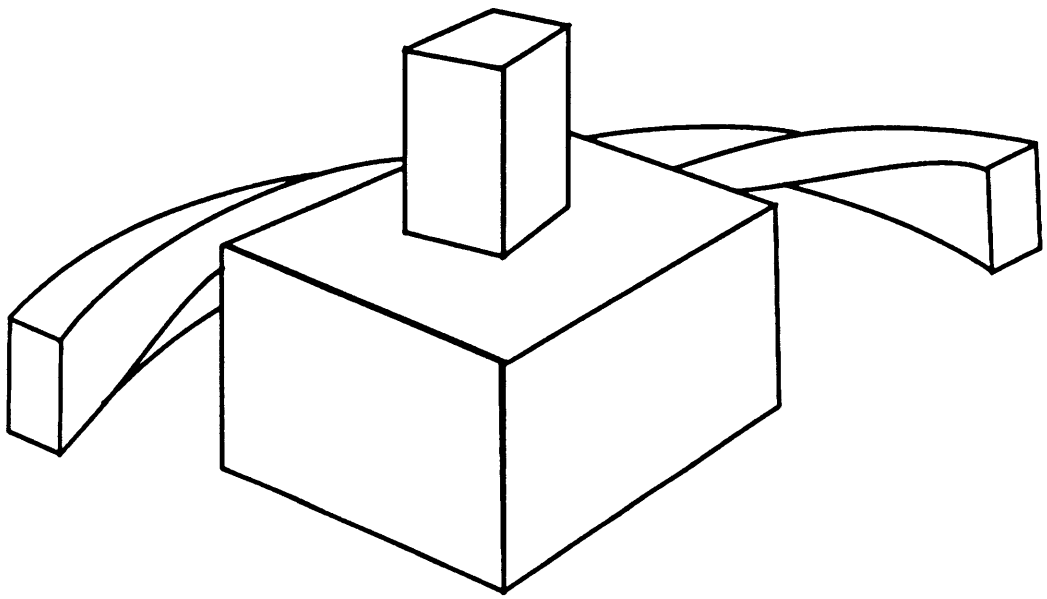


Figure 4 A blocks-world example where the non-accidental property "collinear" is ignored (see text for discussion).

Here we have written the conditions using the probability ratios appearing in the Bayesian formula (2). The constant δ should be chosen such that we consider probabilities larger than $1 - \delta$ as virtually certain in order that the information condition rules out features that simply confirm virtually certain events. Also, in terms of δ, the genericity condition requires that the property P have a probability larger than δ and thus P is not virtually impossible. The particular choice of δ and a quantitative threshold for L are left open in the above proposal. We expect that the choice of these quantities would depend on the utility or risk involved in making, or failing to make, the appropriate inferences, which we do not pursue here. Finally, note the desirability that the inference can be made reasonably often. That is, the context C should not be too rare, and given the generic property P, the

measurements F should also be common. This new requirement has been incorporated as part of the informativeness condition.

2.4 Non-monotonicity of Inferences

We close this section with one final example of the role context plays in our proposal. Most people see Figure 4 as depicting three blocks: one block resting on top of another, and a third twisted block that lies behind. Note that two of the vertical lines associated with the Y-junctions are actually collinear in the image, creating the useful (non-accidental) collinear feature suggested by Lowe (1985). This feature certainly satisfies our likelihood ratio condition. So why don't we see the two blocks as having collinear edges in 3D with one block floating above the other? (A similar example having an accidental view of a "Y" vertex, due to Steve Draper, is given by Hinton (1977).)

To understand the use of collinearity as a feature, we consider inferences appropriate for three different contexts. Each of these contexts is simply a statement about regularities in the scene generating process, and are not meant to imply different stages in the perceiver's visual information processing system. The first context is an "open context", C_{open}, which consists of randomly placed line segments. In particular, collinear, coterminating, or parallel lines in the world are non-generic (i.e. probability zero) in this context. However, although the likelihood ratios for all these properties are easily seen to be large, as was the case for the "V" feature discussed earlier, the a priori probabilities for these "non-accidental" properties are too small to warrant their inference. Hence in the context C_{open} the overwhelmingly probable conclusion is that the collinear, coterminating, and parallel lines in the image simply arise due to some cause other than being the projections of their corresponding 3D properties. (An obvious possibility is measurement noise and a special view of the scene.)

Now consider a second context, C_{group}, similar to the first, but with regularities added that make, say, collinear lines or parallel edges much more probable than they would be in the unstructured context C_{open}. For example, such a context would result if there are processes in the world that cause the 3D line segments or edges to form structures having particular regularities such as textured flow fields (Stevens, 1978; Kass & Witkin, 1988) or blocks with parallel faces (Lowe, 1985). Now the significant prior probability of these specific structures in that context and the large likelihood ratio provided by the non-accidental feature, together ensure that the inference of the corresponding 3D structure is reliable. Given Figure 4 in this context then, and given the alignments and parallel edges, one might infer that these image elements arose from a related group of 3D objects (as indeed they did!).

The third context involves a collection of blocks, C_{block}, where the blocks can rest on one another or float about freely. If blocks float freely then their position and orientation

with respect to the other blocks is assumed to be random, with vanishing a priori probabilities R_{prior} for collinear or parallel edges. So again the situation is analogous to the case of the V-junctions presented earlier (Figure 3). Hence, although the likelihood ratio L is high in context C_{block}, the prior probability that the two blocks would be floating in just such a way to make a pair of edges collinear is vanishingly small, and the resultant a posteriori probabilities R_{post} rule against the interpretation that the two edges happen to be collinear. Instead, we favor some other cause, such as an accidental viewpoint. Finally, we note in passing that the occluded twisted block in Figure 4 is seen as just that – a single block but not as two, although none of the edges are collinear. However, in the context C_{block}, it is reasonable to expect that the implicit axes of the right and left portions of the twisted block could be extracted. Such features satisfy a cocircularity regularity (Parent & Zucker, 1989), which is also a "non-accidental" property, and hence the "one block" inference is justified.

Our point then is that the context in which the scene configuration arose is crucial to the interpretation of a feature, since a change in context can reverse the appropriate inference. In our example, the 3D collinearity conclusion is justified only in the middle context C_{group}; in the less structured context C_{open} and in the most structured context C_{block} the 3D collinear regularity for these lines is not viable. Hence the appropriate inference is non-monotonic with the degree of structure or specification within the context (McCarthy, 1980; McDermott & Doyle, 1980; Reiter, 1980; Salmon, 1967).

3.0 Model Classes

A major point of our analysis of "what makes a good feature" is that supportable inferences are context-sensitive. Features must be evaluated in terms of generic properties or regularities in a specialized context or model class, as contrasted with an open context like a "random-world" model. Implicit in this treatment is that the external world indeed has some non-arbitrary structure, and that our own internal models can express this structure in terms of certain regularities explicitly stated as part of the model. How are these regularities expressed in the Bayesian formalism, and how can they be mirrored in the perceiver's conceptualization of the world?

In an attempt to capture the notion of a regularity, within a probabilistic representational system of a perceiver, Barlow (1985) proposed "good features" should satisfy the "suspicious coincidence" condition $p(A\&B) >> p(A)p(B)$, where A and B are two observations.[2] The intent of the condition is to notice special situations that are not expected by an independence assumption of the occurrence of A and B. Although "suspicious" implies to us that there is a current context, this is not an explicit part of Barlow's proposal, which

[2]Based on the text, we assume that the intended inequality is as appears here. However, note that for the independent event hypothesis, the inequality can be applied in either direction.

requires the very controversial computation of estimating context-free probability distribution functions (i.e. $p(A) = \Sigma p(A|C)p(C)$ summed over all possible contexts). Barlow (1990) discusses at length elsewhere how a neural system might learn the appropriate distribution functions (see also Clark & Yuille, 1990).

One way to capture the intent of Barlow's proposal within the Bayesian framework is to consider the feature observation in the context C_p where the associated property is generic, as contrasted with the current, less specialized context C_o where the property (or properties) are non-generic. More specifically,

> **Suspicious Coincidence**: The observation of a feature F represents a suspicious coincidence in the context C_o if there is a more specialized (i.e. detailed) context C_p such that,
>
> (i) the likelihood ratio involving feature F and property P is large in both contexts, and
>
> (ii) the probability of P in the specialized context C_p is much larger than in the current context C_o, that is $p(P|C_p) >> p(P|C_o)$.

(9)

For example, in our discussion of the blocks in Figure 4 we first considered the open context C_{open} of random lines. The collinearity feature F has a large likelihood in context C_{open}, but the prior probability of 3D collinear lines is negligible. However, in the grouping context C_{group}, the prior probability is significant and the likelihood ratio is still large. Hence, we would consider the observation of collinear lines in context C_{open} as a suspicious coincidence with respect to the more structured context such as C_{group}. Note that this conclusion is not to be considered a reliable inference that context C_{group} actually occurs in the world. (An analysis similar to the one presented in Section 2 could derive suitable additional conditions to ensure a reliable inference of the new context.) Rather, Barlow's notion of suspicious coincidences simply provides an approach for chaining through to more detailed contexts as further regularities are uncovered and assimilated. We do not pursue this chaining process here, and instead concentrate on how a specific context might be represented.

Clearly an internal model can not be expected to match exactly the behavior of external events. In terms of our Bayesian proposal, the internally represented probability density functions $p(P|C_i)$ can not be identical to their external world counterparts, $p(P|C_w)$, say. In particular, as the contexts become more and more specialized (and hence the measures on the probability density functions become more and more biased), the world model and the perceiver's conceptualizations may diverge. We would like to minimize the effects of this divergence. In other words, we seek model contexts, properties, and features that are robust under errors in our estimates of the conditional probability measures. This is a

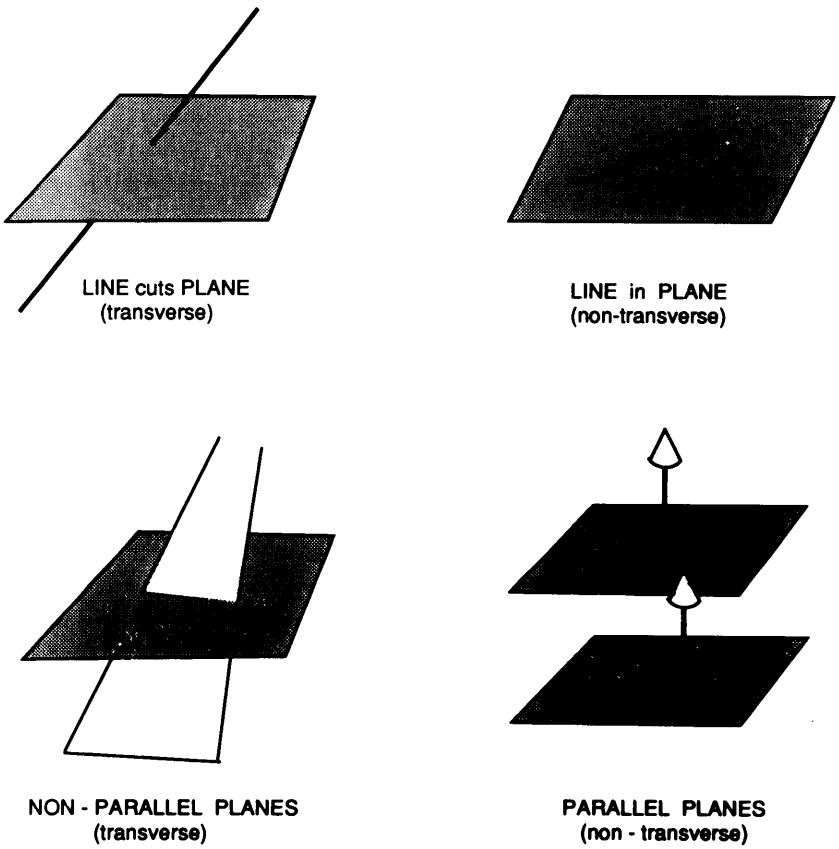


Figure 5 Two kinds of regularities, transverse (left) and non-transverse (right).

lines in 3D is a non-transverse event, but two lines skewed and non-intersecting in 3-space would be a transverse arrangement.

Non-transversality, then, appears at first blush to be the "non-accidental" proposal of Lowe (1985). However, here we use the terminology "transversal and non-transversal" because these terms are context-sensitive and can be applied to world models with arbitrary statistical properties. Thus, in a non-random world model, say one describing body parts, the arrangement such as the V-vertex which we previously considered non-transverse can become transverse (because this is the configuration of an arm). However, in this same model class, the T-junction or parallel line configuration would continue to be non-transverse. Still another example would be an assumed model context where objects are taken to obey two-fold reflectional symmetry. Then a line perpendicular to a plane will be a transversal arrangement, whereas in the absence of such a symmetry constraint, such

a 90 degree intersection is non-transverse. Hence the notion of transversality also involves categorical properties considered special in the current model class. An important type of world regularity can be specified by adding on top of this categorical structure an indication of whether or not a particular non-transversal category has a non-zero prior probability of occuring.

3.2 Key Features

Let us define a model space $\mathcal{M}$ simply as a manifold constructed by parameterizing some modelling domain. The parameters could be involved in descriptions of (3D) position, attitude and shape of various parts, or reflectance properties of surfaces, or higher order structures such as the sounds of a babbling brook. Also various categories P are represented as subsets of the model space, some of which form non-transversal submanifolds within $\mathcal{M}$. For example, our two sticks "V" example corresponds to a model space R^{10}, where the ten parameters describe the position and orientation of the sticks. Consider the category P for which the two sticks form a V-junction (for simplicity, with a particular pair of endpoints). This is a 7-dimensional hyperplane in our model space. We note in passing that this 7-dimensional space has other "special" configurations within it, such as the 5-dimensional hyperplane representing the situations when the two sticks are also collinear.

Next we need to specify how $\mathcal{M}$ and the various categories are meant to represent (or "mirror") structure and events in the world. In particular, we assume a fixed mapping between events in the world and categories within $\mathcal{M}$. The stick example suffices to illustrate the mapping between coterminating sticks in the world, and the representation of this event in $\mathcal{M}$. To avoid unnecessary details we simply identify a world property as P_w, and use P_m to refer to the corresponding category within $\mathcal{M}$. Given this correspondence, we can take a world context C_w (which the reader may assume is simply an index to an appropriate probability density function) along with the associated probability distribution $p(P_w|C_w)$, and consider the "ideal probability distribution" induced on the model space, namely $p(P_m|C_m) = p(P_w|C_w)$. Of course, this ideal probability measure in NOT to be considered part of the perceiver's conceptualization. However, we need to make an assumption about its general structure, namely

Mode Hypothesis: Given a model space $\mathcal{M}$ and a context C_w then the probability measure $p(m|C_w)$ can be decomposed into the sum $\sum_{i=0}^{n} p_i(m|C_w)$ for $m \in \mathcal{M}$. Here p_0 is the background measure and p_i for $i > 0$ is a measure having support only on the non-transversal category P_i within $\mathcal{M}$. Each of these measures is assumed to have density functions of the form

$$p_i(m|C_w) = \mu_i(m)\beta_i exp(-H_i(m|C_w)), i = 0, \ldots, n \tag{10}$$

for $m \in \mathcal{M}$ (see Skilling, 1991). Here μ_0 is the Lebesgue measure on $\mathcal{M}$ and μ_i for $i > 0$ are Lebesgue measures on the property spaces P_i (i.e. delta distributions). The terms β_i can be taken to be 0 or 1, depending on whether the i^{th} mode is a regularity in context C_w. Finally, the remaining terms involving H_i provide a reweighting of the uniform Lebesgue measures; they are exponentiated simply to insure the weights are positive.

The Mode Hypothesis can be seen to be a hypothesis about the form of the "ideal" probability density, for properties within a model class (Bobick, 1987; Marr, 1970). The basic idea is that robust features should supply reliable inferences over a wide range of possible choices for the specific background probability density and for the non-transverse probability densities. In other words, the robustness of the inferences should follow from the structure of the probability density, which in the ideal case will be a collection of delta functions. Ideally, all the perceiver needs to maintain is the locations of these delta functions, but not knowledge of their probability distributions $p(P_w|C_w)$ because typically this information will not be available. Instead we take the (perhaps, extreme) position that an assumed context, C_m, is simply a specification of which categories P_i have a non-zero probability mass. In terms of equation (10), C_m specifies which normalization constants β_i are nonzero, but says nothing about the details of the actual density functions in terms of the weight functions $H_i(m|C_w)$. Different modes can be selected in different contexts, and that is the only control of (assumed) context the perceiver has. For convenience we will abuse the notation, and take $p(m|C_m)$ to mean *any* one of the set of density functions which satisfy equation (10), and is nonzero only on the selected modes specified by the model context C_m.

The stick example provides a concrete case, where the world context consisted of two randomly placed sticks. The particular probability density p_0 is assumed to be a smooth function of both the location and orientation of the two sticks. Such a distribution can be written in the form presented for a background measure. Many different choices for H_0 are possible, describing for example a uniform distribution within a cube, or a Gaussian distribution, etc. The important property of p_0 is that, *independent* of the choice of H_0,

it assigns zero probability to all non-transversal manifolds such as the P_i of $\mathcal{M}$. Suppose there are two regularities in this particular world context. One causes the two sticks to form a V-junction with a non-zero probability, and the other causes these V-junctions to form the degenerate case of collinear sticks. Such a world satisfies the Mode Hypothesis, with the V-junctions and the collinear V-junctions forming the only non-transversal sets which have positive probability mass. Within this particular context, such regularities will support robust inferences from their measurements, even though the (unavailable) density functions associated with the perceiver's internal model space C_m do not match exactly the associated objective density functions in the world, namely $p(P_w|C_w)$.

To support this claim, we now proceed to develop the relation between the special class of non-transverse properties $P_i \in \mathcal{M}$ and their associated features F_i. Hence, in addition to a model space $\mathcal{M}$, we now require a measurement space $\mathcal{I}$ and an imaging mapping, π, from $\mathcal{M}$ onto $\mathcal{I}$. (This basic set up is similar to that used in Observer Mechanics (Bennett et al., 1989) with the exception that for us the various spaces and mappings are all part of the perceiver's representational framework. For Observer Mechanics these entities *are* the world.) Features F_i are identified with subsets or submanifolds within the measurement space $\mathcal{I}$. To illustrate this mapping, consider again the two stick case. Then, given orthographic imaging, the 10-dimensional configuration space for two sticks will be imaged to a 6-dimensional feature space. Within this feature space is the 4-dimensional hyperplane (a non-transversal set) consisting of all possible images containing V-intersections. We assume that the imaging map π correctly models the qualitative structure of the transduction and subsequent measurement processes of the perceiver (again, detailed noise models are not assumed). Finally, we define the probability of a feature F, say $p(F|P\&C_w)$ to be the probability induced by the image map and the measure on $\mathcal{M}$. That is, $p(F|P\&C_w)$ is given by the probability of the set of all models m which image to F, namely $\pi^{-1}(F)$. Similarly, given a model context, $p(F|P\&C_m)$ is taken to mean any one of the induced measures consistent with the model context C_m.

A model class is defined to be a pair of spaces $\mathcal{M}$, $\mathcal{I}$, along with the imaging map π. In addition to these spaces a model class includes two lists of categories, one a list of model properties (or categories) P_i within $\mathcal{M}$, the other a list of features F_i within $\mathcal{I}$. Finally, a particular model context C_m for a perceiver is simply a selection, from the list of categories P_i, of those which are assumed to have a non-zero mass in the "ideal" probability measure. Given this framework, we obtain our robust feature:

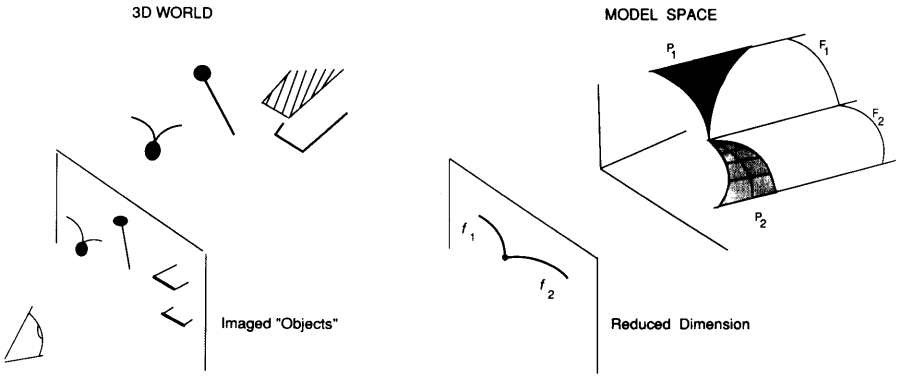


Figure 6 Two different event spaces to which our proposal (11) applies: Left, 3D "objects" projected into the image plane; Right, a high order space parameterized by α, β, γ with features F_1 and F_2 that provide constraint surfaces for these parameters (or constraint lines, f_1 and f_2 in the case of the reduced dimension).

3.3 A Simplified Internal Model

To begin, we assume that the perceiver has the ability to parse events in any given model space into configurations consisting of points, line segments, edges and corners (Figure 7). In this first example, we also assume that the events in the 3D world are similarly points, lines, edges and corners, which are imaged onto a 2D space. (Later, we will consider world events that are not these simple geometric primitives and more general types of features.) In order to recognize non-transversal arrangements of these primitives, the perceiver must also have available concepts that help define the "interesting" relations between them. These concepts act like axioms in geometry or number theory. They dictate the fundamental nature of the world as we see it. Here, we choose notions of coincidence, parallelness, perpendicular, collinear and coplanarity.[3] It is understood that the perceiver understands that

[3]Clearly there are other choices, such as cocircularity, special tesselations, etc. Just which concepts are selected is of course a critical issue, but beyond the scope of this paper.

SIMPLIFIED INTERNAL MODEL

A. **OBJECTS** in the model space are constructed from Points, Lines (Segments) and Planes (Facets).

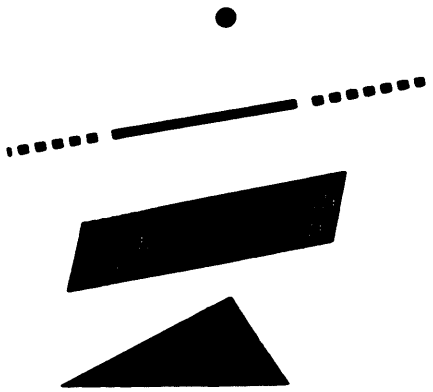
B. **OBJECT ELEMENTS**

Point

Line Segment (Bar)

Edge (of Region)

Corner (Facet)



C. **CONCEPTS** (innately) available to the perceiver.

1. "Object " Type: point, line, segment, etc.
2. "Object Relations: parallel, coincident, perpendicular, collinear, co-planar (symmetry).
3. "Special" Property: gravity.

D. **CONTEXT** (or model class)

Variable over contexts.

Figure 7 The basic ingredients of the observer's internal model.

POINT TO LINE SEGMENT

CONCEPT	DEPICTION	COST	CODIMENSION 3D	CODIMENSION 2D
COINCIDENT (end)		α, β, γ	3	2
COLLINEAR (on)				
		α, β	2	1
(off)				
PERPENDICULAR		γ	1	0
CO-PLANAR		0	0	0
PARALLEL	– undefined –		N/A	N/A

Figure 8 Non-transverse arrangements of a point to a line segment.

LINE TO LINE SEGMENT

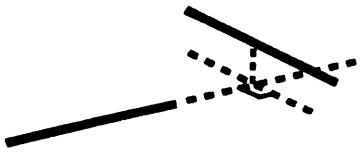
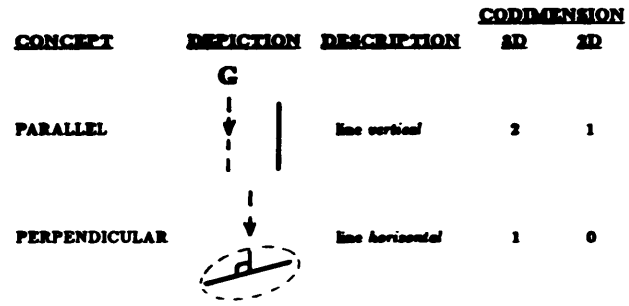
CONCEPT	DEPICTION	CODIMENSION	
		3D	2D
COINCIDENT		5	3
COLLINEAR		4	2
PERPENDICULAR			
(non-planar)		1	0
(co-planar)		2	0
PARALLEL		2	1

Figure 9 Non-transverse arrangements of one line segment to another, again in a "random world" context.

LINE AND GRAVITY



POINT TO LINE SEGMENT (PLUS GRAVITY)

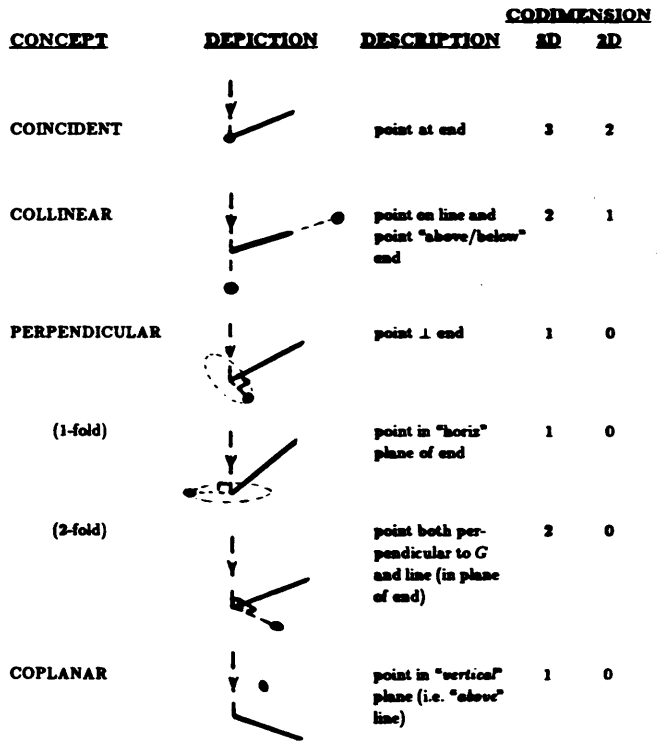


Figure 10 The addition of a coordinate frame, such as the gravity vector, expands the Key Feature possibilities.

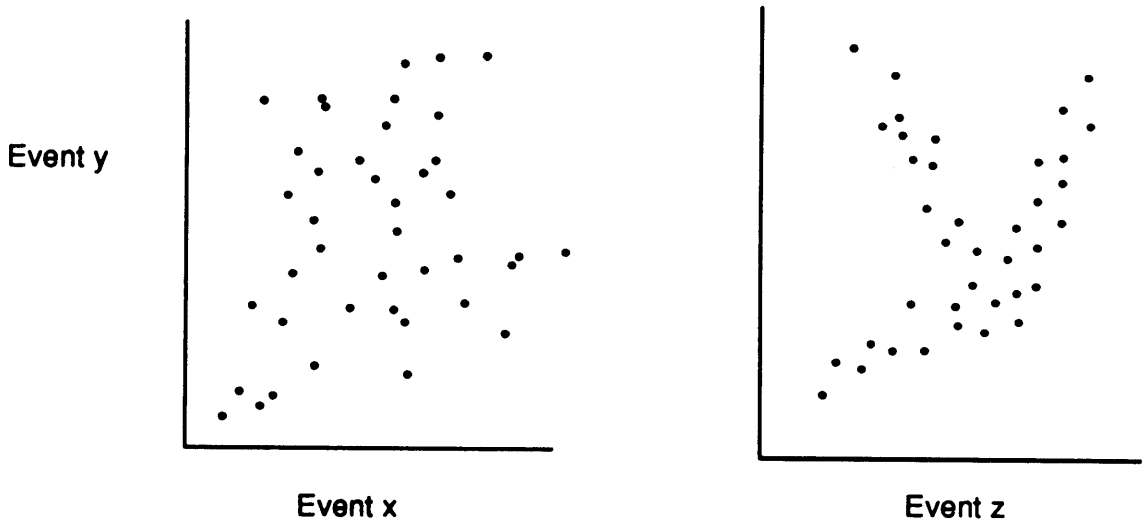


Figure 11 Left: A cluster (or perhaps two!) of points whose specialness is difficult to demonstrate statistically. Right: A pattern of points that is much simpler to show is non-arbitrary, not only because the subspace is more coherent, but especially because the arrangement is non-transversal for a simple line-segment model.

field). And, finally, the measurements on the image will be noisy. Hence, we can expect to see distributions of points in the event spaces, not well-marked trajectories. Clearly a random cluster of points, such as Figure 11a (left) can not support a key feature, whereas Figure 11b (right) looks promising. How then do we proceed to test whether the observed distribution of points in the event space supports a key feature? Fortunately, a good part of the necessary machinery is available, provided that one knows in advance the possible model types that apply (Kendall, 1989). But this is indeed the case because all the "low-order" types of Key Features have been enumerated. The procedure, then, is simply to test the hypothesis that the points in the feature space support one of the Key Feature configurations known to the perceiver.[5]

4.1 Data Description

To illustrate a version of Shape Statistics, consider the configuration in Figure 11b. We know that the coincidence of three lines is a special configuration of codimension 2 in the

[5]Note that Kendall & Kendall (1980) provide a very detailed analysis of the collinear Key Feature applied to the data of Stonehenge in order to test the hypothesis that the alignments marked some interesting astronomical event.

event space. The task is then to obtain a probability density function (*pdf*) for each line and separately for their intersection. To estimate each line (and hence its trajectory), we can create a density function concentrated along a 1D curve or spine, following the methods of Leclerc (1989) or Hinton et al. (1991). Denote this spine together with its associated *pdf* as a "caterpillar". An important property of these approaches is that such caterpillars provide an appropriate form of description for each "image". In particular, for Figure 11b we might expect that a process similar to Leclerc's would extract a description in terms of three straight caterpillars. Their width would be determined from the scatter of the data points perpendicular to the spine. In addition, the endpoints of the linear segments would also be provided only to within the same resolution. Similarly, for 11a, the same process might be expected to choose a description involving only one or two blobs.

Given these descriptions it is now clear how to deal with images such as Figure 11b. Presumably we have recovered precisely three line segments along with an estimate for possible errors in the positions of the endpoints. This provides a "stick image", to which we can apply our usual repertoire of Key Feature models (i.e. candidate configurations). The only difference is that we have an explicit estimate for the noise variability, so we could expect to get more detailed estimates of the basic probabilities and likelihoods in our Bayesian proposal.

It is interesting to note the similarity in our proposal for good model descriptions and good features. For example, the "three stick" configuration is a specialization of a description including polynomial spines, suggesting that lower dimensional descriptive models can be found on particular *nontransversal* submanifolds in higher dimensional descriptive spaces. The observation that an interpretation is close to one of these non-transversal sets suggests that we collapse the description to the smaller space. This is analogous to observing a non-transversal feature in our model class.

4.2 Decision Rules

The extraction of a good description for Figure 11b, followed by the inference of a triple junction, is clear in principle but it raises some difficult issues. Both Figures 11a and 11b are fairly clear cut in terms of their structure, with only one model fitting very well in either situation. However, consider adding more noise to Figure 11b to obtain some intermediate cases. Presumably the parse into three separate lines becomes less certain, as does the quantitative data on the parameters for the lines. In an abstract feature space the picture is of a noise estimate associated with each feature which covers a larger region as the input noise is increased. A final point is that, in terms of our Bayesian proposal, the likelihood ratio L for observing particular regularity will decrease (basically, by adjusting the width of the caterpillar we are keeping $p(F|P\&C)$ roughly constant, but this increased width will also cause the probability of false targets, $p(F|notP\&C)$ to increase). As a

result the inferences will become less certain or, once the Informativeness Condition fails, uninformative.

We discussed the problem of choosing a good description of the data in the previous section. Given a description we are now faced with choosing an appropriate inference from our model class. How can such a decision be made? Simple structural rules, such as choosing the most singular model (highest codimension) consistent with the data description, or the least singular model, can easily be shown to be inappropriate. Similarly, the maximum likelihood description will generically be a transversal point in the feature space, and thus the regularities will almost never be inferred. Recall that the regularities only support strong inferences if their *a posteriori* probabilities are sufficiently large, and the likelihood ratio L for features associated with properties serves as the amplification factor from a priori probabilities to a posteriori probability ratios. A decision rule based on maximum a posteriori probability (MAP) estimates is possible, given estimates for the prior probabilities (Clark & Yuille, 1990). However, it is not clear that such useful estimates on the priors are possible to simply memorize, especially when we need these priors for each of a wide range of contexts. Thus for MAP estimation to work we need to estimate the priors on the fly from the model class, with the one glimmer of hope here being that the estimates may only need to be accurate to within an order of magnitude, or so. A different approach involves placing a partial order on various possible interpretations (see Jepson and Richards, 1991, 1992). This partial order could be made on the basis of probability estimates, or some other form of preferance relation. For example, for the blocks in Figure 4 we may estimate that a floating collinear interpretation (codimension 4) is significantly less probable than an accidental view interpretation (codimension 1 or 2 depending on whether or not the blocks are assumed to be right angled), especially since we have no way of explaining this codimension 4 event. Difficult research issues remain for the resolution of these problems.

4.3 Ideal Observers

Recently, Bennett, Hoffman & Prakash (1989) have constructed a probabilistic framework called "Observer Mechanics" which provides an alternative model for both the world and the perceiver. The major component of this model is an "observer" which is the 6-tuple (X, Y, E, S, π, η) where (loosely speaking) X is a configuration space of quantities being observed, and Y is the imaging space formed by the many-to-one mapping $\pi : X \rightarrow Y$. Within X lies a set E of "distinguished configurations" that play the role of our non-transversal categories. The images of configurations within E form the set of features S observed in Y. Hence S corresponds to our non-transversal image features. Finally, for each $s \in S$, $\eta(s, \cdot)$ is a probability measure on $\pi^{-1}(s)$.

An ideal observer is defined in terms of an unbiased measure μ_x on the configuration space X. We take this measure to be the probability of a particular configuration in X, but in the absence of any structuring influence producing the distinguished configurations

captured in E. That is, μ_x is analogous to our background probability distribution p_0. Within this framework, an observer is then said to be ideal if

$$\mu_x[\pi^{-1}(s) - E] = 0.$$

In other words, when there is no regularity or structure in E, there is a zero probability of observing an element of S that does not result from an element of E (i.e. the probability of a false target, is zero). In terms of our earlier example, the probability of a "V" image feature is just the probability of the set of all configurations in X which project to S, namely $\mu_x(\pi^{-1}(S))$. In a random stick world this probability is zero, and this implies that the previous equation must be satisfied (see the discussion around equation (3.3) in *Observer Mechanics*). Therefore, there exists an "ideal observer" for 3D "V"'s in a random stick world. In fact, if we identify the set E with world property P and identify the set $S = \pi(E)$ with image feature F, then $F = \pi(P)$ using our terminology an ideal observer can be constructed precisely when:

> **Ideal Observer Proposal**: The image feature F is non-generic in the absence of world property P, and occurs with probability 1 in the presence of world property P.

Besides the condition that F occurs with probability one in the presence of P (which may be regarded as a consequence of our definition of $F = \pi(P)$), the only condition on an ideal observer is that the false target rate must be zero. Hence the measurement likelihood ratio must be infinite. Thus ideal observers are similar to our key features, in that both require an infinite likelihood ratio L. However, unlike key features, ideal observers include situations such as the "V" observer in a random stick world, even in the absence of a world regularity for "V"'s. In addition, ideal observers include the case of two randomly placed sticks, where the world property P is simply the occurence of non-parallel sticks. This property occurs with probability one, yet there is still a feature having an infinite likelihood ratio. In our Bayesian proposal we include conditions that eliminate cases such as these. In particular, the V-observer is eliminated by the requirement that the world property is generic, and the skewed-sticks observer is eliminated by the informativeness condition.

Observer mechanics recognizes this problem but deals with these degenerate cases in a rather different manner. Both the V-observer and the skewed-sticks observer are essentially "no-op" observers. The V-observer in a random stick world detects a feature with probability zero, so it never reports a V observation. On the other hand, the skewed-stick observer detects its feature with probability one, and always responds. In both cases, the performance has zero probability of being wrong, which justifies the term "ideal". The conclusions of these "no-op" observers can reliably be used as input to other observers,

and that is the primary requirement on an ideal observer. The problem we posed in this paper is different, we actually want useful, robust, and informative features. As a result, our definition of a key feature is (roughly) a subset of the situations for which there is an ideal observer, and to specify this subset we require structure both in the regularities of the world and in the conceptualization of the perceiver.

A second difference between our formulation and observer theory is that given a feature, we attempt to make categorical statements about world properties within a model context, whereas observer theory strives to place probability measures on world properties that are supported by observing a particular feature. Given a feature s, the conclusion of the observer is provided by a probability measure $\eta(s, e)$, with e in the distinguished space E (corresponding to P). This measure $\eta(s, \cdot)$ is called the interpretation kernel. In our framework this distribution is the a posteriori probability distribution $p(m|F\&P)$, conditional on both the feature F and the property P. For example, given the skewed-stick observer, the interpretation kernel would provide the a posteriori probability for the 3D position and orientation of the two sticks. In contrast, our approach provides only the categorical response that the two sticks are indeed skewed in 3D. The computation of such a interpretation kernel clearly involves detailed a priori probability distributions, which we have attempted to avoid. However we note that, in situations where the priors can be computed, the incorporation of analogs to the interpretation kernal could play a role in extending our "categorical" good feature formulation. For our purposes in this paper, we only point out that the most plausible approaches for the computation of these priors involve the manipulation of assumed regularities in the world, which again ties in with our notion of a model class.

5.0 Examples

Our treatment of Key Features within a feature space has been limited to configurations built from points, lines, edges, and facets. Although we have tried to stress that these elemental object types are not the only primitives that one might use, it is easy to regard our treatment as applying only to a "blocks world". The essential point, however, is that it really doesn't matter what sensory attributes or dimensions we consider, nor the particular object types chosen as "observable" primitives in that space of features. For example, we could explore non-transverse configurations in time rather than space, or frequency-time as in an acoustic feature space (Bregman, 1990). Here, however, we will present three further examples taken from vision.

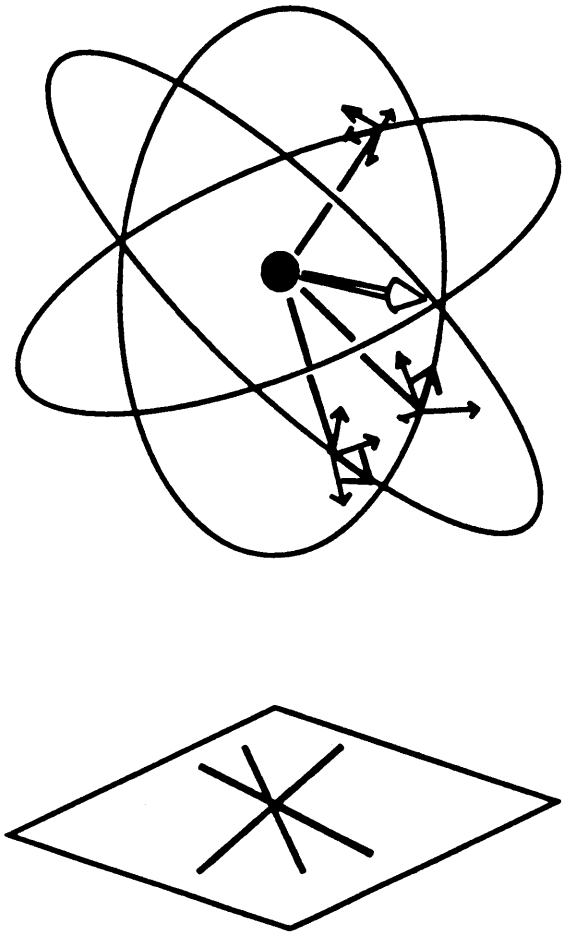


Figure 12 A Key Feature for the translation direction for ego motion has the same type of non-transversal configuration as that for finding the spectral quality of the illuminant!

addition, the power of the key feature might be further augmented if we also have extrinsic frame vectors that act like the gravity vector in Figure 12, such as those derived from vestibular inputs. This space housing the key feature for Ego Motion is thus much like that shown earlier in Figure 11 (right) where events in the feature space lie on loci that radiated from a single vertex. Here, then, we have a specific instance where noise and resolution will affect the robustness of the key feature.

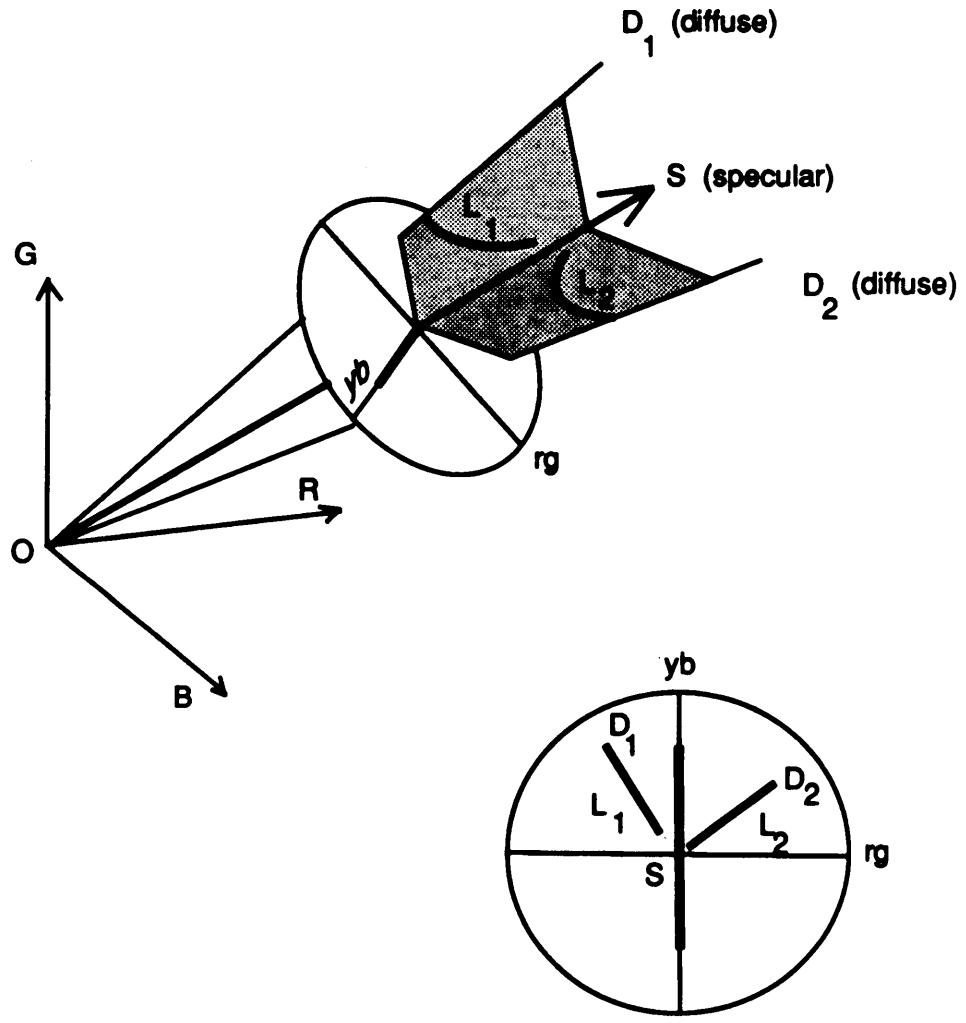


Figure 13 Top: Representation in the (R, G, B) space of responses L_1 and L_2 to two surface patches, lit by the same source S, that have different diffuse components of reflectance (D_1 and D_2). The two planes described by L_1 and L_2 intersect along the axis S, which describes the chromaticity of the illuminant, because the specular component of reflectance is common to both objects. The responses from two or more objects that define distinct planes can thus be used to find the axis S that describes the chromaticity of the illuminant. Bottom: Projection of L_1 and L_2 onto the chromaticity plane $rg - yb$. The lines described by the responses intersect at the point S marking the chromaticity of the illuminant. If the perceiver's model incorporates the knowledge that most daylight illuminants lie on a segment of the yb axis, as indicated, then two patches suffice to define a "crow's foot" key feature configuration. (Adapted from D'Zmura & Lennie, 1986.)

spaces and their projections are quite consistent with our proposal, and would appear to be physiologically plausible. However, note that in such mappings that mirror particular "real world" properties, the co-dimension of a key feature becomes ambiguous and, as mentioned earlier, it is the inferred property that is assigned the codimension associated with the particular key feature configuration observed internally.

6.0 Summary

Previously, others such as Binford (1981), Lowe (1985), and Witkin and Tennenbaum (1983), have noted that good features should reflect "non-accidental" configurations that are specially informative yet typical of the world (such as two parallel lines). However, we note that the intuitively robust character of an inference based on a non-accidental feature is not simply due to the fact that they have a large likelihood ratio (i.e. the feature is expected when the world property is present, but very rare in the absence of the property). In the discussion of our Bayesian Proposal we have shown that a large likelihood ratio is clearly not sufficient to ensure robust inferences (see also Knill & Kersten, 1991). Rather, the likelihood ratio simply serves as a lever for raising the *a priori* probability of the particular world property. Given too low an *a priori* probability this lever is insufficient to provide a high a posteriori probability and hence a robust inference. This notion of a reasonably large prior probability is implicit in the discussion of a non-accidental feature, and explicit in the presentation of the intuition behind Observer Theory, yet the full impact it has on the definition of a good feature was not made explicit.

The analysis of the two block example in Figure 4 shows that the definition of a good feature must include a specification of the cognitive context in which it is being used. The collinearity feature, a classic non-accidental feature, is reliable in some contexts but nonsense in others. The difference hinges on what the perceiver is willing to assume are regularities in the world. Thus good features are necessarily bound to the current context of analysis, to conceptual models, and to the regularities that a perceiver expects to be operative (MacKay, 1978, 1985). The fact that a feature can be good in one context, but nonsense in a more specialized context, reflects a common phenomena in inductive inference known as non-monotinicity (Salmon, 1967). Whether your bias is for perceivers who maintain a detailed probabilistic model of their world, or for those which use a logical framework, this non-monotonic behaviour must be dealt with by the explicit use of contextual information (McDermott & Doyle, 1980; Reiter, 1980).

Given that the specification of "good features" requires the specification of the current context, we suggest a model class as an appropriate form for representing contextual information. Basically a model class is an abstract space of models about the world, which has been carved up into various categories. Some of the categories are transversal, representing open subsets of the space. Other categories exist on subsets (submanifolds) of the parameter space and have a smaller dimension than that of the embedding space. These latter

categories are non-transversal, and their degree of specialization can be roughly measured by their codimension, that is, the difference in dimension between the embedding space and the particular category. In addition, the model space can be projected to the image, where a similar categorization in terms of transversal and non-transversal image features can be made. Our canonical example is of a non-accidental property or feature such as collinear lines, which is non-transversal in both the world and image spaces. Indeed we pursue our proposal in some detail for such geometric features, but we also show it has applications to other domains such as motion or colour interpretation.

So far this conceptualization is independent of whether or not certain categories support robust inferences in that it does not specify whether any non-transversal category reflects a regularity in our world. There is no notion of probabilities in this categorization. To fully specify a model class we need to select particular categories as corresponding to regularities that are considered possible within the current context, thus entertaining Bayesian-like propositions (Pearl, 1990). However, we prefer to keep the categorical conceptualization itself independent of the notion of regularities, or of probabilities in the world, to allow for the same set of categories to be used in a host of different contexts. Given the regularities, a Key Feature supports the inference of a particular non-transversal but generic world category (i.e. one expected or selected by the pereiver). Hence such a feature carries within itself its appropriate interpretation, in that the regularity has already been specified in the world, and this step of the inference process becomes rather trivial. Finally, given the appropriate qualifications provided by the Bayesian Proposal, such a key feature can be expected to provide a reliable inference for that particular regularity in the world.

For a structured, non-arbitrary world and for a defined set of (internal) concepts about primitive object types and their possible relations, the set of Key Features can be enumerated. All such features are not equally powerful with respect to their inference strength. As a measure of this power, we suggest the codimension of the Key Feature configuration, with respect to the class of models computable in the feature space. Our proposal requires a slightly different view of "feature detectors" than that customarily taken. Rather than simply providing a "measurement" as an oriented bar mask might do, our "feature detector" recognizes a non-transverse configuration in an event space constructed from such measurements. The class of configurations recognizable are only those non-transverse arrangements that can be computed for the types of object primitives and relations specified. The principal task, then, is to discover the object types used to construct the event spaces, for these will generate the model classes. We suspect that the relations computed within the different event spaces will be similar, and relatively trivial. Their reliability, of course, will depend upon how well the conceptual relations and primitives match the actual building blocks and constraints imposed by Nature on constructions in the real world.

References

Airy, H. (1870) On a distinct form of transient hemiopsia. *Phil. Trans. Roy. Soc. Lond.*, **160**: 247-264.

Barlow, H. (1953) Sensation and inhibition in the frog's retina. *J. Physiol.*, **119**: 69-88.

Barlow, H. (1961) Possible principles underlying the transformations of sensory messages. Chapt. 13 in W. Rosenblith (Ed.), *Sensory Communication*, Cambridge, MA: MIT Press. (See also pp. 782-786.)

Barlow, H. (1985) Cerebral cortex as model builder. Chapt. 4 in D. Rose & G. Dobson (Eds.), *Models of Visual Cortex*, NY: Wiley.

Barlow, H. (1990) Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Res.*, **30**: 1561-1571.

Bennett, B., Hoffman, D. & Prakash, C. (1989) *Observer Mechanics.* London: Academic Press.

Binford, T.O. (1981) Inferring surfaces from images. *Art. Intell.*, **17**: 205-244.

Bobick, A. (1987) Natural categorization. MIT Art. Intell. Lab. Tech. Report 1001.

Bregman, A.S. (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge, MA: MIT Press.

Clark, J.J. & Yuille, A.L. (1990) *Data fusion for sensory information processing systems.* Boston: Kluwer.

Curtis, S.R. & Oppenheim, A.V. (1989) Reconstruction of multidimensional signals from zero crossings. Chapt. 5 in S. Ullman & W. Richards (Eds.), *Image Understanding*, Norwood, NJ: Ablex.

Desimone, R., Albright, T.D., Gross, C.G. & Bruce, C. (1984) Stimulus selective properties of inferior temporal neurons in the Macaque. *Jrl. Neurosci.*, **4**: 2051-2062.

D'Zmura, M. & Lennie, P. (1986) Mechanisms of color constancy. *Jrl. Opt. Soc. Am. A*, **3**: 1662-1672.

Feldman, J. (1991) Perceptual simplicity and modes of structural generation. *Proc. 13th Ann. Conf. Cog. Sci.*, August.

Feldman, J. (1992) Constructing perceptual categories. *Proc. Comp. Vis. & Pat. Recog.*, to appear.

Frost, B.J. (1985) Neural mechanisms for detecting object motion and figure-ground boundaries contrasted with self-motion detecting systems. In D. Ingle, M. Jeanerod & D. Lee (Eds.), *Brain Mechanisms of Spatial Vision*, The Netherlands: Nijhoff.

Gershon, R., Jepson, A.D. & Tsotsos, J.K. (1987) Highlight identification using chromatic information. *Proc. IEEE 1st International Conference on Computer Vision, London, England*, pp. 161-170.

Gibson, J.J. (1950) *The Perception of the Visual World.* Boston: Houghton Mifflin.

Gross, C.G., Desimone, R., Albright, T.D. & Schwartz, E.L. (1985) Inferior temporal cortex and pattern recognition. *Pont. Acad. Sci. Scripta Varia*, **54**: 179-201. The Vatican.

Hinton, G.E. (1977) *Relaxation and its role in vision.* Ph.D. thesis, University of Edinburgh.

Hinton, G.E.,Williams, C.K.I. & Revow, M.D. (1991) Adaptive elastic models for hand-printed character recognition. In *Advances of Neural Information Processing Systems 3.*

Hubel, & Wiesel, T.N. (1959) Receptive fields of single neurons in the cat's striate cotex. *J. Physiol.*, **149**: 574-591.

Jepson, A.D., Gershon, R. & Hallett, P.E. (1987) Cones, color constancy, and photons. In J.J. Kulikowski, C.M. Dickinson & I.J. Murray (Eds.), *Seeing Contour and Color*, Oxford: Pergamon Press.

Jepson, A. & Heeger, D. (1990) A fast subspace algorithm for recovering rigid motion. *Proc. IEEE Workshop on Visual Motion, Princeton, NJ*, pp. 124-131.

Jepson, A. & Richards, W. (1991) What is a percept? MIT Cog. Sci. Memo #43.

Jepson, A. & Richards, W. (1992) A lattice framework for integrating vision modules. *IEEE-Sys. Man. & Cyb.*, in press.

Kass, M. & Witkin, A. (1988) Analyzing oriented patterns. Chapt. 18 in W. Richards (Ed.), *Natural Computation*, Cambridge, MA: MIT Press.

Kendall, D.G. (1989) A survey of the statistical theory of shape. *Statistical Sci.*, **4**: 87-120.

Kendall, D.G. & Kendall, W.S. (1980) Alignments in two-dimensional random sets of points. *Adv. Appl. Prob.*, **12**: 380-424.

Knill, D.C. & Kersten, D.K. (1991) Ideal perceptual observers for computation, psychophysics and neural networks. In R.J. Watt (Ed.), *Pattern Recognition by Man and Machine*, London: MacMillan.

Koenderink, J.J. & van Doorn, A.J. (1981) Exterospecific component of the motion parallax field. *Jrl. Opt. Soc. Am. A*, **71**: 953-957.

Leclerc, Y.G. (1989) Constructing simple stable descriptions for image partitioning. *IJCV*, 3(1): pp 73-102.

Lee, H.-C. (1986) Method for computing the scene illuminant chromaticity from specular highlights. *Jrl. Opt. Soc. Am. A*, **3**: 1694-1699.

Lee, H.-C. (1989) A computational model of human color encoding. Eastman Kodak Report, Imaging Science Laboratory.

Lettvin, J., Maturana, H.R., McCullock, W.S. & Pitts, W.U. (1959) What the frog's eye tells the frog's brain. *Proc. IRE*, **47**: 1940-1951.

Lowe, D. (1985) *Perceptual Organization and Visual Recognition.* Boston: Klewer.

Marr D. (1970) A theory of cerebral neocortex. *Proc. R. Soc. Lond. B.*, **176**: 161-234.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco: Freeman.

McCarthy, J. (1980) Circumscription – a form of non-monotonic reasoning. *Art. Intell.*, **13**: 27-39.

McDermott, D. & Doyle, J. (1980) Non-monotonic logic. *Art. Intell.*, **13**: 41-72.

MacKay, D.M. (1978) The dynamics of perception. In P.A. Buser & A. Rougeul-Buser (Eds.), *Cerebral Correlates of Conscious Experience*, Amsterdam: Elsevier.

MacKay, D.M. (1985) The significance of 'feature sensitivity'. Chapt. 5 in D. Rose & V.G. Dobson (Eds.), *Models of the Visual Cortex*, New York: Wiley.

Parent, P. & Zucker, S.W. (1989) Trace inference, curvature consistency and curve detection. *IEEE Trans. on Pat. Anal. & Mach. Intell.*, **11**: 823-839.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, CA: Morgan Kaufman.

Perrett, D.I., Rolls, E.T. & Caan, W. (1982) Visual neurons responsive to faces in the monkey temporal cortex. *Exp. Brain Res.*, **47**: 329-342.

Poston, T. & Stewart, I. (1981) *Catastrophe Theory and Its Applications.* London: Pitman.

Regan, D.M. & Beverley, K.I. (1982) How do we avoid confounding the direction we are looking and the direction we are moving? *Science*, **215**: 194-196.

Reiter, R. (1980) A logic for default reasoning. *Art. Intell.*, **13**: 81-132.

Reuman, S. & Hoffman, D. (1986) Regularities of nature: the interpretation of visual motion. In A. Pentland (Ed.), *From Pixels to Predicates*, Norwood, NJ: Ablex.

Richards, W. (1975) Visual space perception. Chapt. 10 in E.C. Carterette & M.P. Friedman (Eds.), *Handbook of Perception, Vol. V: Seeing*, New York: Academic.

Rose, D. & Dobson, V.G. (1985) *Models of the Visual Cortex.* New York: Wiley.

Salmon, W. (1967) *The Foundations of Scientific Inference.* Pittsburg: Univ. Pitts. Press.

Shafer, S.A. (1984) Using color to separate reflection components. *Color Res. & Appl.*, **10**: 210-218.

Skilling, J. (1991) Fundamentals of maximum entropy in data analysis. Chaptr. 2 in B. Buck, V. Macaulay (Eds.), *Maximum Entropy in Action*, London: Oxford University Press.

Sober, E. (1975) *Simplicity*. London: Oxford University Press.

Stevens, K. (1978) Computation of locally parallel structure. *Biol. Cybernetics*, **29**: 19-28.

Tinbergen, N. (1951) *The Study of Instinct*. Oxford: Clarendon.

Thorpe, W.H. (1963) *Learning and Instinct in Animals*. Second edition. Cambridge, MA: Harvard University Press.

Witkin, A. & Tennenbaum, J.M. (1983) On the role of structure in vision. In J. Beck, B. Hope & A. Rosenfeld (Eds.), *Human and Machine Vision*, New York: Academic.

Zeevi, Y.Y. & Shamai, S. (1989) Image representation by reference-signal crossings. Chapt. 6 in S. Ullman & W. Richards (Eds.), *Image Understanding*, Norwood, NJ: Ablex.