Spatial Vision in Humans and Robots

The Proceedings of the 1991 York Conference on Spatial Vision in Humans and Robots

Edited by

Laurence Harris
York University
North York, Ontario

Michael Jenkin
York University
North York, Ontario

CAMBRIDGE
UNIVERSITY PRESS

What makes a good feature?

Allan D. Jepson
Whitman Richards

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.

1. Introduction

In 1870 Lord Airy noted that human visual processing made special use of oriented line segments. His inference was based upon the fortification pattern observed during a migraine attack. Roughly one hundred years later Hubel & Wiesel (1959) confirmed this inference by direct recordings of neurons in the visual cortex of mammals. Since then, there has been a tremendous surge in the discovery of other neurons in all sensory modalities that are optimally sensitive to some specific feature of the sensory array (Rose and Dobson, 1985). In vision these range from the low-order space-time derivatives of intensity such as moving "edge" or "line" detectors, to

more complicated patterns that include various symmetries, such as faces or hands (Barlow, 1953; Albright et al., 1984b; Gross et al., 1985; Lettvin et al., 1959; Perrett et al., 1982). To some extent, we expect these observed features to reflect the demands of survival imposed upon the species. Thus the high-level features found in primates are not expected to occur also in simpler animals, such as the fish or frog. Across species, therefore, we find an enormous spectrum of features, especially if we include those specialized trigger patterns or "innate releasing mechanisms" reported by the ethologists (Thorpe, 1963; Tinbergen, 1951). Given this vast collection, it might seem unlikely that one could abstract away some principles that define "what makes a good feature?" However, here we attempt to do just that.

Our guiding hypothesis is that "seeing" is the inference of world properties from image elements—i.e. the various patterns of intensities on the retina. A "feature" is typically viewed as a measurement of image structure, at the level for example of Marr's primal sketch (Marr, 1982). Clearly, many different kinds of measurements or "features" are possible. Intuitively, however, those most often sought after will point directly and reliably to a unique, *meaningful* event in the world. But the criterion that a feature be meaningful implies that the perceiver has some goal or context in mind. For example, for a baby gull the significance of a red spot in the image depends on whether it is seen in the context of a traffic light or as coloration on the beak of an adult gull (Figure 1). In the context of a beak, its salience is sufficient to trigger a feeding response. Somehow the gull is primed to immediately make the necessary inference. Hence we propose that "what makes a good feature" should include the property of having a ready explanation for its appearance (MacKay, 1978; MacKay, 1985).

Under this view, a simple intensity change, an oriented image-edge, or a "zero-crossing" segment analyzed in an open context is not a very good feature. Although we know that edges contain the bulk of information in an image (Barlow, 1961b; Curtis and Oppenheim, 1989; Zeevi and Shamai, 1989), many factors can create the intensity changes that trigger the "edge detector". These include shadows, material changes, scratches, occlusions, etc. Hence there is no unique structure that can be induced from a single intensity edge or line. Consequently, although line elements or edges may be our initial primitives, by themselves they do not exhibit structure over which useful inductions can immediately be made.

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

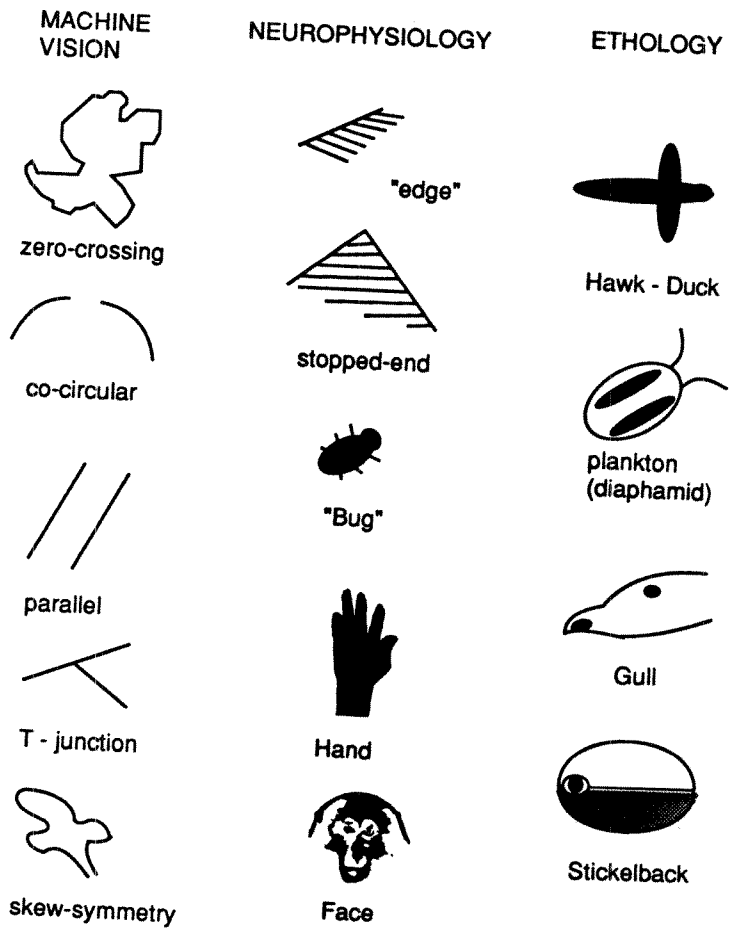


Fig. 1. *Typical features proposed by machine vision, neurophysiology, and ethology. What common properties do these features satisfy? What makes one feature better than another?*

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 al., 1989; Marr, 1970; Reuman and Hoffman, 1986; Witkin and 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.

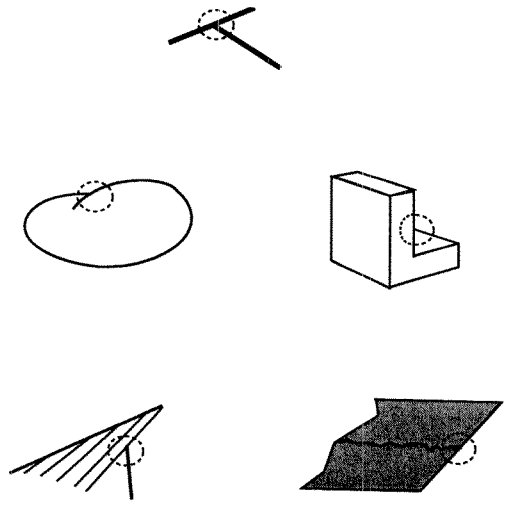


Fig. 2. *If the image primitives are contours (such as zero crossings), then features typically can be created in many ways. For example, the T-junction may arise either from an occlusion or from an actual T-vertex in 3D. Hence the interpretation associated with a feature depends strongly on the context. 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.*

2. 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))

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

and the likelihood ratio L is defined to be

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

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

$$L = \frac{p(F|P\&C)}{p(F|notP\&C)} >> 1 \tag{5}$$

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.

2.2. The importance of significant priors

To illustrate how the reliability of an inference depends on both a large likelihood ratio and a generic world property, consider a context consisting of a random 3D arrangement of two sticks. In this context consider the *non-generic* property that two sticks form a "V" intersection in 3D. For our analysis it is more convenient to let property P include both perfect and nearly perfect "V"'s (see Figure 3). That is, for some tolerance ϵ, property P means the endpoints of two sticks come within the distance ϵ of each other in 3D. The measurement provided in the feature F is simply that the projected distance (with respect to some specified ray) between the two endpoints is less than ϵ. This feature formally consists of all stick configurations in which both endpoints lie somewhere within the depicted cylinder. (Informally F consists of two endpoints lying within a disc of radius ϵ in an "image" plane formed by orthographic projection). The measurement F holds when the sticks have property P, and therefore $p(F|P\&C) = 1$. On the other hand, if the two sticks do not satisfy property P then the probability of F is simply proportional to the area of the disc of size ϵ. The likelihood ratio L is therefore proportional to $1/\epsilon^2$, which is much larger than one for small values of ϵ. Hence this situation satisfies the measurement likelihood ratio condition.

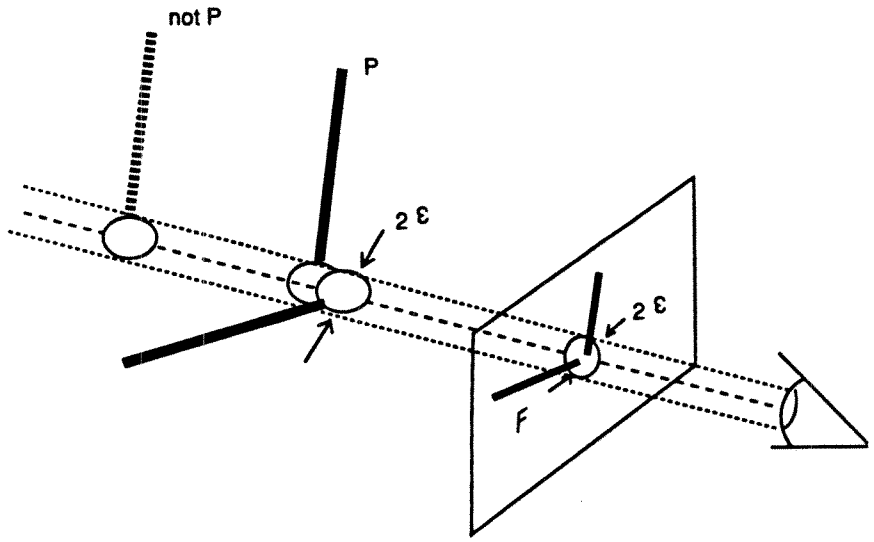


Fig. 3. *Two sticks in 3D form a near-V vertex to create property P, which projects into the V-junction image feature F. The resolution for the sticks forming a V is taken as a disc of radius ϵ in the image (assuming orthographic projection) and, for the 3D tolerance, the sphere of similar radius. Although 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.*

Given that $L >> 1$, should we then infer the 3D property "V", given the measurement F of such a V-junction feature? Surprisingly, in our chosen random world such a conclusion is almost always guaranteed to be *WRONG.* Hence, in this model context, the feature F could hardly be a worse indication of the intended world property P. The probability of an endpoint lying within the sphere to form a "V" is much lower than the probability that the endpoint lies anywhere in the cylinder. More specifically, the joint probability of property P and having feature F, in our model world, is proportional to the volume of the ball of radius ϵ around the endpoint of a stick. That is, $p(P\&F)$ is proportional to ϵ^3. However, feature F is satisfied whenever the endpoint of the second stick lies anywhere within a cylinder of radius ϵ about the ray passing through the endpoint of the first stick and hence $p(F)$ is proportional to ϵ^2. Consequently the conditional probability of property P, given the feature F and a random world context will be $O(\epsilon^3/\epsilon^2) = O(\epsilon)$. Thus for small ϵ we are almost guaranteed to be wrong if we infer P. The

appropriate inference would be to infer "*notP*", that is, the endpoints do not actually come close in 3D.[1]

Referring again to the decomposition of R_{post} (equation (2)), we see that the problem with our "V" example is that even though the likelihood ratio L provided by F is proportional to $1/\epsilon^2$ and is thus much larger than one, it is not large enough to amplify the prior probability ratio $R_{prior} = 0(\epsilon^3)$ to a reasonable level. To correct this, we need a significant prior probability that an endpoint lies within a particular sphere, i.e. $R_{prior} > \delta$. In 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 and 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 the probability of this feature is $p(F|P\&C) = 1$. However, since the orthographic image of two parallel lines must also be parallel, it follows that $p(F|notP\&C) = 0$. Therefore the image feature F consisting of two skewed lines provides an infinite likelihood ratio for L. Also the genericity condition is satisfied since the sticks have both a random position and orientation. Therefore it follows that the probability of such a skewed arrangement in the random world is one and R_{prior} is infinite. Hence R_{post} must also be infinite, and the inference is certain. Nevertheless, such a feature is simply confirming the obvious and should not be included in our definition of a good feature. This can be corrected by adding a condition that the *a priori* probability $p(P|C)$ is not too close to one. (An analogous situation also occurs when a property P is so overwhelmingly unlikely that, even after the observation of F, the *a posteriori* probabilities favor *notP*. This case is caught by the requirement of significant priors discussed above.)

[1] This problem was discussed at length some years ago at a workshop on Perceptual Organization arranged in 1984 by A. Pentland and A. Witkin. See also Knill and Kersten (1991) for another example.

Hence to insure that a feature not confirm the obvious, we add the following condition:

Informativeness Condition: Given a context C and a constant $\delta > 0$, the property P occurs with probability $p(P|C) < (1 - \delta)$, or, equivalently,

$$R_{prior} = \frac{p(P|C)}{p(notP|C)} < \frac{1-\delta}{\delta} \tag{7}$$

Collecting our conditions together, we now arrive at the following proposal for a good feature:

Bayesian Proposal: Given a constant δ, a good image feature F for world property P in (world) context C satisfies

1 Likelihood ratio condition:

$$L >> 1/\delta;$$

2 Genericity condition:

$$R_{prior} = p(P|C)/p(notP|C) > \delta/(1-\delta)$$

3 Informativeness condition:

$$R_{prior} = p(P|C)/p(notP|C) < (1-\delta)/\delta,$$

and $p(F|P\&C)$ and $p(C)$ are significantly bigger than zero.

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

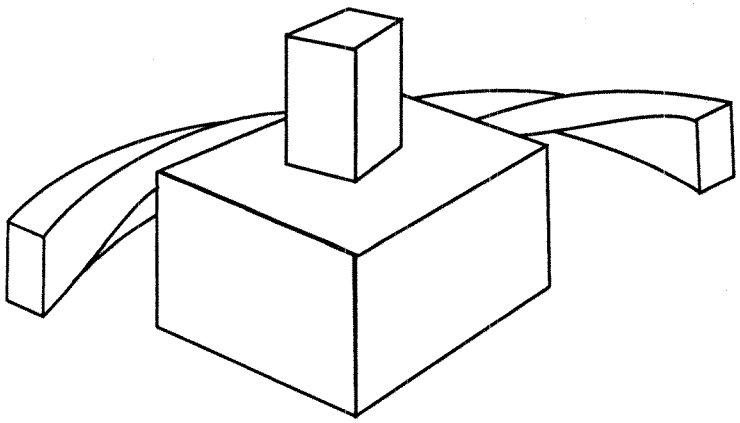


Fig. 4. *A blocks-world example where the non-accidental property "collinear" is ignored (see text for discussion).*

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 and 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 and 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 and Doyle, 1980; Reiter, 1980; Salmon, 1967).

3. 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 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 and 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,

1 the likelihood ratio involving feature F and property P is large in both contexts, and

2 the probability of P in the specialized context C_p is much larger than in the current context C_o, that is

$$p(P|C_p) >> p(P|C_o).$$

For example, in our discussion of the blocks in of 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

[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.

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 different type of robustness than was considered in earlier, where the appropriate probability distributions and the particular context were given as facts.

One class of properties in which robustness is (nearly) ensured in the face of modeling errors are those that are "non-accidental", such as the collinearity of two sticks. First we consider such properties as idealizations where our resolution ability is unlimited; later we return to the issue of dealing with finite resolution. More specifically, we assume here that the likelihood ratio for the collinearity feature, for example, goes to infinity as the measurement error, ϵ, decreases to zero. If we consider a world context, C_w, which has a positive probability mass for situations in which two 3D sticks are precisely collinear, then the prior probability ratio R_{prior} is also at least as large as this positive constant. Equation (2) then shows that the *a posteriori* probability ratio R_{post}, and thus the reliability of the inference of collinear lines in the world, can be made arbitrarily large by taking ϵ sufficiently small. For our earlier non-accidental feature the "V", a similar idealization is to put a point mass of probability (i.e. a Dirac distribution) at the occurrence of the 3D "V" intersection (see Figure 3). Then, as we make ϵ smaller the contribution of this point mass stays fixed, while the probability of the remainder of the cylinder reduces to zero. As a result, the *a posteriori* probability ratio of a "V" intersection goes to infinity as ϵ decreases, and the correct inference is virtually assured. Note that this is a constraint on the shape of the probability density function, rather than on its detailed value. The following describes a sub-class of regularities that meets this condition.

3.1. Two kinds of regularities

Given any model for objects or properties in a world, the structural regularities associated with that model can be divided simply into two classes: those configurations or relations that arise when the elements of the model

are positioned arbitrarily with respect to one another, and those that require special placements (Poston and Stewart, 1981). For example, let our objects be a line and a plane, and let our assumed model of structural relations to be nil—in other words there are no specialized arrangements in the world. Then if the line and plane are each thrown out haphazardly in 3-space, we expect the line to intersect the plane at some arbitrary angle (Figure 5). An alternate configuration, such as the line lying exactly in the plane, is impossible, unless someone placed it there. These two configurations depict respectively transversal and non-transversal intersections of two objects. Intuitively, the notion of transversality is one of event stability between objects: slight perturbations of the arrangement do not affect the topology. For example, a knife plunged into an apple would create a transversal cut (unless precisely radial), whereas the cut would be non-transversal if the knife were tangent to the apple as if peeling its skin. In the latter case, proper peeling requires precise alignment with the tangent plane of the apple. Such events which are not stable to slight perturbations of the elements that create it are called "non-transversal". Thus, given an assumed context of random stick-world, the "V" vertex formed by the two 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 occurring.

3.2. Key features

Let us define a model space $\mathcal{M}$ simply as a manifold constructed by parameterizing some modeling 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

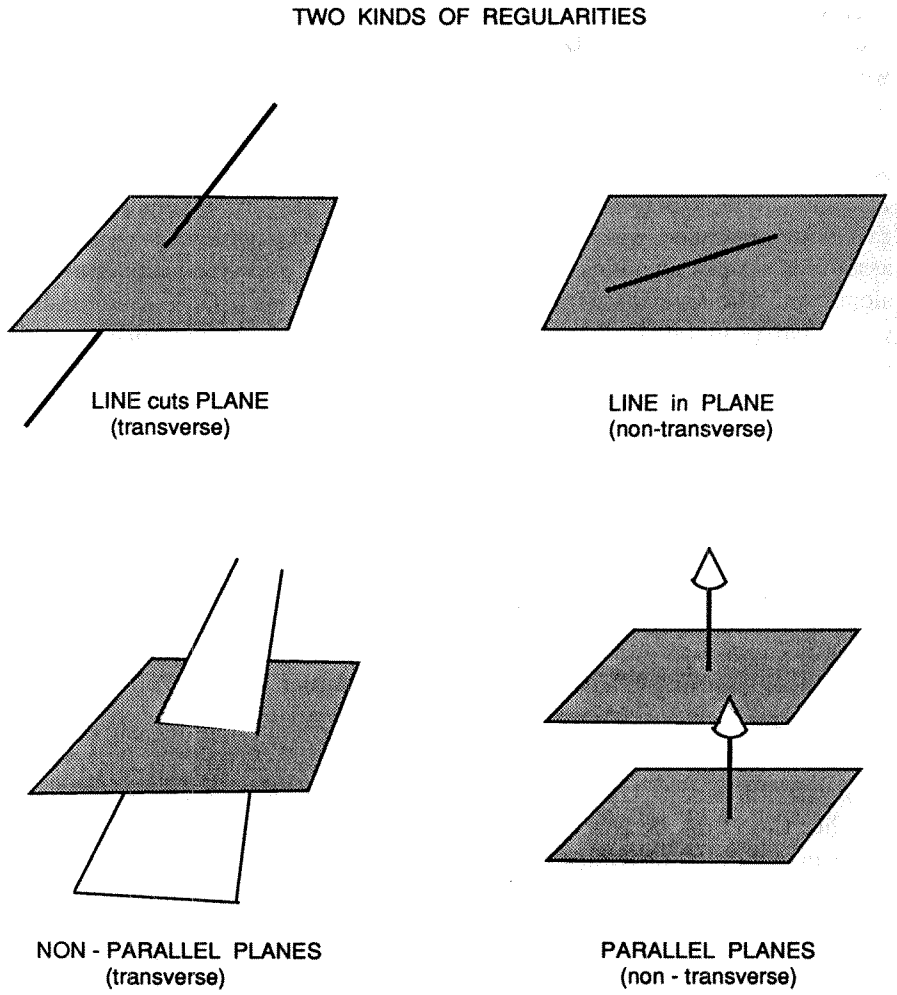


Fig. 5. *Two kinds of regularities, transverse (left) and non-transverse (right).*

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

$$p_i(m|C_w) = \mu_i(m)\beta_i exp(-H_i(m|C_w)), i = 0, \ldots, n \qquad (8)$$

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 (8), 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 density functions which satisfy equation (8), and is non-zero 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:

Key Feature Definition: Given a model class and a model context C_m, then F is a Key Feature for a world property P in context C_m if

1 P is non-transverse within $\mathcal{M}$, yet generic in C_m (i.e. $p(P|C_m) > 0$);

2 the probability of the feature in the absence of property P is zero,

$$\text{(i.e. } p(F|notP\&C_m) = 0); \tag{9}$$

3 $p(F|P\&C_m)$ is greater than zero.

More simply put, F is a key feature for a property P if P is a generic nontransverse mode in the model space, and F occurs in the presence of P, but never in its absence. (Hence in the special case where F is non-transverse and generic, F will be a key feature provided that the conditions for P are satisfied.) Notice that we only refer to zero and positive probabilities in the Key Feature Definition, which is appropriate because no particular positive values are specified by the model context C_m. The fact that F is a key feature is independent of the detailed quantitative structure of the "ideal" probability measure $p(P_w|C_w)$. Rather, as desired, it depends only on the proper selection of active modes. Hence what becomes critical is not just the types of measurements used to construct a model space, but rather the types of submanifolds within this space that the perceiver can recognize or build (see Feldman, 1991, 1992, and Sober, 1975, for additional constraints on such submanifolds). For example, in Figure 6 on the left we see three configurations projected onto an image plane, which can be directly viewed as key-feature arrangements for a "point" and "line" in a random world, or for two planes. On Figure 6 (right) however, we envision a model space having parameters α, β, γ, which contains various property categories, e.g. P_1 and P_2. The features involve measurements that result in constraint surfaces within the model space (e.g. F_1 and F_2 for the full space, or f_1 and f_2 for the reduced space). Concrete examples of such constraint surfaces for observer motion or the inference of surface reflectance are given in Section 5. For now, we simply point out that the structure of the intersection of two such constraint surfaces can provide a "key feature" for a world property.

By using the earlier formalism provided in Section 2 one can easily check that key features provide a robust inference of world property P_w.

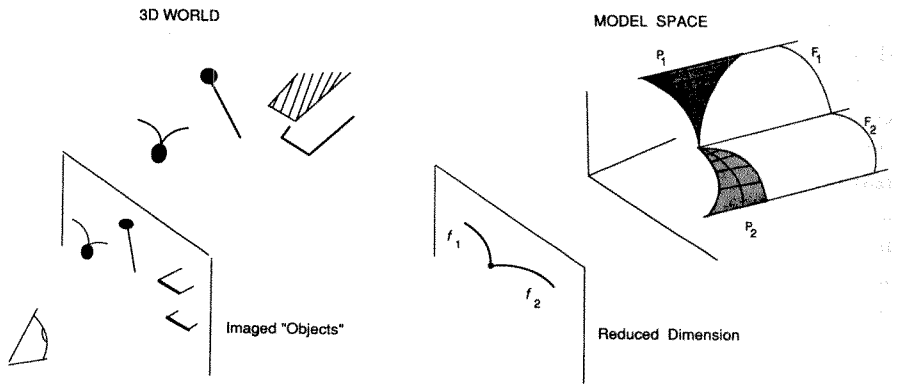


Fig. 6. *Two different event spaces to which our proposal (9) applies: Left, 3D "objects" projected into the image plane; Right, a high order space parameterized by α, β, γ with features F_1 and F_2 that provide constraint surfaces for these parameters (or constraint lines, f_1 and f_2 in the case of the reduced dimension).*

Specifically, referring to the definition (9), the last two requirements ensure that the likelihood ratio is infinite, while the first ensures that the *a priori* probability must be positive. Therefore the *a posteriori* probability ratio for property P is infinite, that is, P is certain to occur. Clearly, a key-feature must be an idealization to have such strong properties. Indeed we have assumed that there is no resolution limit on the measurement process, and that the modes themselves have no *transverse* structural variability (i.e. our ideal modes are non-transverse, whereas noise can be expected to make them transverse). Shortly we will deal with these issues. First, however, we show how, given any model space, the special class "Key Features" can be identified, and how a measure can be placed on the "informativeness" of each feature in that class.

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 Peano's axioms in geometry. 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 coincidence applies to points, end-points of lines, planes, etc. and recognizes the distinctions between these types of coincidences. Similarly, intrinsic to the concept of parallelness is the knowledge that this relation applies only to lines or planes, etc. (These conditions can be formalized, but the formalization adds little to the understanding of our proposal.) Finally, we allow knowledge of "special" concepts that may be defined outside the particular model space, but can be mapped into it. The gravity vector for a "blocks" world model space would be an example. These concepts, then, define the perceiver's internal model for the property space under consideration.

3.4. Key feature enumeration

Given a well-formulated internal model, it is a relatively straightforward task to enumerate the form that the different key features will take. In particular, we seek non-transversal properties which image to non-transversal features. The non-transversality of the feature is sufficient to ensure that the feature occurs with probability zero in the absence of all regularities. We begin with point-to-line arrangements, then consider line-to-line, and finally line-to-virtual line, namely the gravity vector. Along the way, a measure of the inferential power of any given feature will also be specified. From these examples, it should be clear how additional key features can be enumerated for any well-specified model class. We consider only the case where our given object relations are generic in our model class, which is taken to be a specialization of an open class. Moreover, we consider measurements consisting of the position and orientation of points and line segments in an orthographic image. In terms of our formalization the features arising from such measurements are constraint sets in the model space. However, since the mapping from these image measurements to the constraint set is fairly intuitive, we ignore this step and consider "features" to be the usual image measurements of position and orientation.

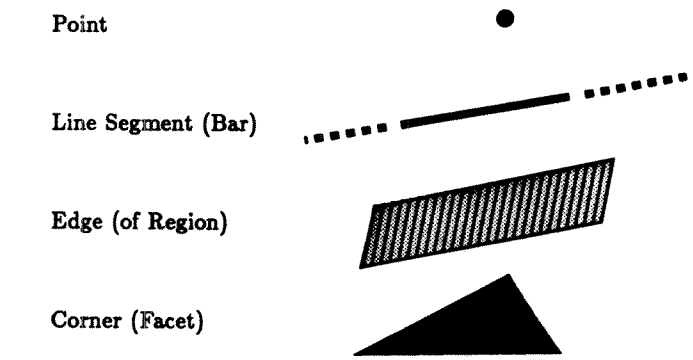
Point to line Consider first the possible non-transverse relations between a point and a line. Let the line be taken as a reference in the 3-space in which the point and line appear. The positioning of the point then has three degrees of freedom, say α, β, γ (corresponding to x, y, z in some coordinate frame built upon the line). For our given internal model, we can entertain only four ways of positioning the point with respect to the line: it can be either coincident (with an end point), collinear (on the line), perpendicular or coplanar—parallel is undefined for this pair. These relations are given in the left column of Figure 8.

[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



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.

Fig. 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

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

For *coincidence*, there is only one relation, namely with the point positioned at the end of the line segment. In 3-space, such positioning costs us three degrees of freedom (DOF) as indicated in the third column: namely, α, β and γ all are now fixed. We call this cost the "codimension" of the arrangement (see Poston and Stewart, 1981), which in this case is *three*. Similarly, in the 2D image the coincidence relation requires that both image coordinates of the point are fixed, giving a codimension of *two*.

The next relation is *collinear*, with the point lying on the line itself or on its extension. Only one degree of freedom (DOF) of positioning remains,

LINE TO LINE SEGMENT

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

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

hence the 3D codimension for this non-transverse arrangement is two (or one in the image plane because the point must lie on an infinite 2D line).

The relation of *perpendicular* between a point and a line occurs when the point lies in the plane of the end point of the line — as if the line were the normal vector to the plane. Because the point can be placed anywhere in this plane and still satisfy the relation, the codimension of this configuration is one. In the image, however, the point can lie anywhere with respect to

the line. Thus the special arrangement is lost and the codimension of this set of images is zero. Finally, the point may be placed arbitrarily in space, creating the single transverse arrangement in 3D. Then a plane is defined, but because the configuration is transverse, the codimension is zero.

The possible key-features given these concepts are limited to just those situations in which the configuration has a positive codimension in both 3D and 2D. That is, for a point and a line only the coincident and collinear configurations qualify.

Line to line In a similar manner we can enumerate all non-transverse arrangements between two line segments, given this particular "model world" (Figure 9). Now, however, we have increased the degrees of freedom for the positioning and pose from three to five. The extra two provide the orientation of one line relative to the other. Hence when two lines are coincident (i.e. collinear with coincident endpoints) there are no remaining degrees of freedom and the codimension is five. Similarly, a *collinear* arrangement between two lines in 3-space will have codimension 4, because one translation remains allowable. For the *perpendicular* configuration we have two cases, one where the two lines are coplanar (codimension 2) and the other when they are not (codimension 1). Finally, two lines can be *parallel*, and this arrangement has codimension 2 because only the orientation (two DOF) of one line to the other is restricted.

The above codimensions were specified for the configurations in a 3-space. However often the observer has available only a projection of the property space. In this case, just as when the three-dimensional world is imaged onto our retina, the specialness of a configuration may be lost. The most obvious instance is when the non-transverse arrangement is specified by an angle, such as "perpendicular". Unless the viewpoint is special, angular relationships are not preserved on projection. In our line-to-line example, the two perpendicular configurations in 3-space will project into arbitrary, transverse relations in the lower dimensional "image" space. Hence these arrangements have codimension zero in 2D, as indicated in the last column of Figure 9. Similarly, we note that coincident lines now have codimension 3 in a plane, collinearity 2 and parallelness 1. Hence only these three latter configurations have the chance of meeting the criteria for a Key Feature.[4]

In our "modal" world where properties are configured to satisfy point, line or planar constructions in a property space, clearly many non-transversal arrangements are possible. The point-to-point and line-to-line examples only illustrate the simplest. If we were to continue to complete the pairwise cases, four more pairs would have to be considered (point-to-point, point-to-face, line-to-face, face-to-face), for five possible relations, not including the two-fold pairing of relations. Multiplying these thirty pairwise cases by the number of added elements for triples, quadruples, etc., rapidly explodes the possibilities. Hence even for our simple model world the space of key features is very rich. Nevertheless, in principle all the non-transversal cases in both the model and image spaces can be identified.

[4] All projections into a lower subspace need not reduce the codimension of the arrangement. For example, a point at the end of a line segment, in the context that the dot must be on the line, is codimension one both in the 3D world as well as in the 2D image. Symmetry constraints are also often preserved in the projection without reducing the codimension.

Line to gravity Often, factors extrinsic to the feature space may impose special frames within the space, as part of the internal model. Such frames create special categories and alter the codimension of events in that feature space. The gravity vector G is a typical example. Because this particular vector defines a virtual line, having no ends nor specific position, the conceptual relations "coincident and collinear" are undefined, and only the parallel relation can be used with this vector. The codimension of a line parallel to the gravity vector is two. Given this special frame vector G, we now have the key feature "vertical line", of codimension one in the 2D image plane, assuming again that lines can be placed arbitrarily.

One final point. Note that if an extrinsic frame such as gravity is imposed upon a feature space, then additional natural concepts such as "vertical" or "horizontal" may be defined within the model class. Further examples are given in the enumeration of point to line segment in a gravity frame (Figure 10, lower).

4. Statistical variability

Although the intuitions behind our notion of a key feature seem compelling, it is useful to consider how our proposal can be extended to natural environments that include both structural variations within a particular world category as well as imaging and measurement noise. As presented, the important notion of non-transversality is an idealization. In practice, the red spot on the gull's beak (or the stickleback's eye) will not lie precisely at a vertex. Or, the process creating a line may leave only point traces (such as a texture flow 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 can not support a key feature, whereas Figure 11b 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

LINE AND GRAVITY

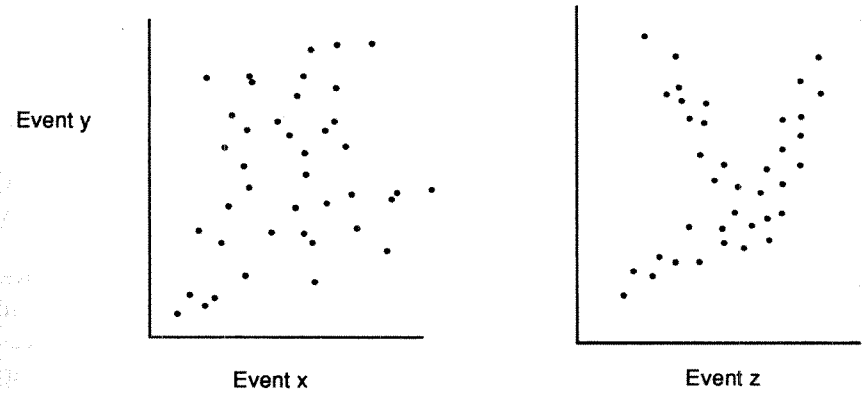
CONCEPT	DEPICTION	DESCRIPTION	CODIMENSION 3D	CODIMENSION 2D
	G			
PARALLEL		line *vertical*	2	1
PERPENDICULAR		line *horizontal*	1	0

POINT TO LINE SEGMENT (PLUS GRAVITY)

CONCEPT	DEPICTION	DESCRIPTION	CODIMENSION 3D	CODIMENSION 2D
COINCIDENT		point at end	3	2
COLLINEAR		point on line and point "above/below" end	2	1
PERPENDICULAR		point ⊥ end	1	0
(1-fold)		point in "horiz" plane of end	1	0
(2-fold)		point both perpendicular to *G* and line (in plane of end)	2	0
COPLANAR		point in "*vertical*" plane (i.e. "*above*" line)	1	0

New Concepts : "*VERTICAL*", "*ABOVE*"/"*BELOW*", "*HORIZONTAL*"

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



(a) (b)

Fig. 11. *(a): A cluster (or perhaps two!) of points whose specialness is difficult to demonstrate statistically. (b): 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.*

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 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. (1992). 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.

[5] Note that Kendall and 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.

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 and 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 preference 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 et al. (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 $\pi : X \to Y$. 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

$$\mu_x[\pi^{-1}(s) - E] = 0. \tag{10}$$

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 occurrence 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 kernel 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. 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.

5.1. Innate releasing mechanisms

Lorentz, Tinbergen and other ethologists (Thorpe, 1963) have noted that certain species-specific stimuli will trigger patterns of behavior in animals. Several examples were illustrated earlier in Figure 1. For example, a red spot near the tip of the beak of an adult gull will elicit feeding behavior from the young chick. Indeed, any such red spot located near the apex of a cone suffices. Clearly this can be idealized as a very non-transverse arrangement. In a 3D world, a spot at the vertex of a cone would have a minimum codimension 3 — even if the cone were given in the class of feature elements. Depending upon the sophistication of the gull's color system, we could easily add another 2 for the codimension of the specific color "red".

Furthermore, this releasing stimulus is generic (all gull parents have the spot) and is modal (there is no sea of red spots near the ends of cones visible to the chick). A similar analysis applies to the red belly of the stickleback, with the eye lying at the vertex formed by the color contour and the front face of the fish (see bottom illustration in Figure 1). The other patterns in Figure 1 also are idealized non-transverse and generic for species. A "stick version " of the hawk-duck configuration in 3-space has codimension 3 when stationary, but codimension 5 when moving along the long axis — the latter projecting into an image event of codimension 3 even if the symmetric relation is ignored. Similarly, the symmetries of the plankton eaten by perch have a high codimension in a world where line elements would otherwise be arbitrary or "fish-like". All of these "events" satisfy the key-feature constraints, and project into significant non-transversal, yet generic configurations with a robust codimension.

5.2. Ego motion

When we move in the world, we effortlessly compute our direction of translation. Only in the case where our fixation direction is aligned with the direction of body motion is the computation relatively simple, for then the optic array has two-fold symmetry as noted by Gibson (1950). However when we look to the side as we move forward, then the optical flow field is complex with gross asymmetries (Koenderink and van Doorn, 1981; Regan and Beverley, 1982; Richards, 1975). Nevertheless, a simple Key Feature can be derived from this flow field (see Chapter 4 and Jepson and Heeger, 1990 and see relevant neurophysiology by Frost, 1985).

Its form is as illustrated in Figure 12. The depiction places the observer at the center of a unit sphere. The flow pattern is on the surface of this field. For each local patch of flow (assuming an arbitrary angle between the direction of translation and the line of regard) there will be a residual, net flow vector. This vector defines a great circle on the unit sphere. The direction of body translation lies on this circle. Because two great circles always intersect (at two points), we need to inspect a third patch of flow to create a triple intersection. This is equivalent to three lines intersecting in a plane and hence has codimension one, which can be potentially increased as more patches are examined. In 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 11b 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 (see Chapter 4).

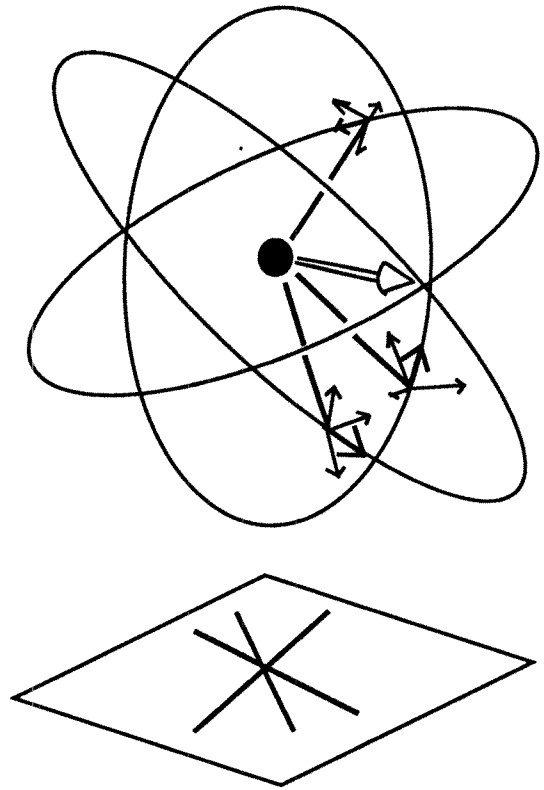


Fig. 12. *A Key Feature for the translation direction for ego motion has the same type of non-transversal configuration as that for finding the spectral quality of the illuminant!*

5.3. Color: finding the spectral content of the illuminant

Our previous examples stressed geometric relations in the world. To make the point that the relevant geometry is not in the world, but rather in the perceiver's representation of the world property, we provide one final example of a key feature.

A classical problem in vision is computing the spectral content of the scene illuminant. This computation is needed in order that this component of the reflectance function can be "discounted" when recovering the spectral reflectance of a surface. Shafer (1984) and Lee (1986) independently proposed a simple model for reflecting surfaces that solves this problem (see also Gershon et al., 1987). The image intensities arising from light reflected off a surface is broken into a matte (diffuse) and specular (mirror)

component. These components add, with weights depending upon the surface orientation, the viewer's position and the reflectance properties of the surface. However, because the model is linear, for any given patch of surface, the locus of observed image intensities in the three color channels must lie on a plane containing the perfect diffuser, the perfect mirror, and the origin (see Figure 13). (We are assuming the ambient light has the same spectral density as the illuminant.) Any two such planes intersect along a line passing through the origin. A third plane that intersects the other two along the same line provides a key feature for the illuminant direction having codimension one. In Figure 13 we have shown a projection of these planes along the intersection line and see that a "Y" vertex is created. Obviously, the strength of the feature can be increased by examining more patches, with each additional patch adding another unit to the codimension.[6]

5.4. Abstract model spaces

Note that both the color and motion Key Features have the same form in their separate spaces. This similarity is important, because it illustrates that the "models" used in any event space can be quite simple, yet still have very significant inductive power, across a range of world properties. Also, in these particular model spaces, the chosen parameterizations seem compelling—matching our intuitions as to which properties might be represented. However, the representation of events within the model space need not be simple lines or planes. Indeed, for complex objects like an animal's face or a tree we should expect that the properties and relations might appear in the form of more complex, curved surfaces which themselves may be "viewed" or projected onto several different lower dimensional spaces to facilitate indexing, for example, as illustrated earlier in Figure 6. Such model 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. Summary

Previously, others such as Binford (1981), Lowe (1985), and Witkin and Tennenbaum (1983), have noted that good features should reflect

[6] The Key Feature may have more structure than that described. For example, if the projection plane is chosen properly, then the lines will be straight. This requires a check on collinearity between samples taken from the same patch. In addition, we have knowledge about where "natural" illuminants should lie in our color space, namely along the black body locus. (This locus acts like the "gravity" vector.) (Jepson et al., 1987; Lee, 1989)

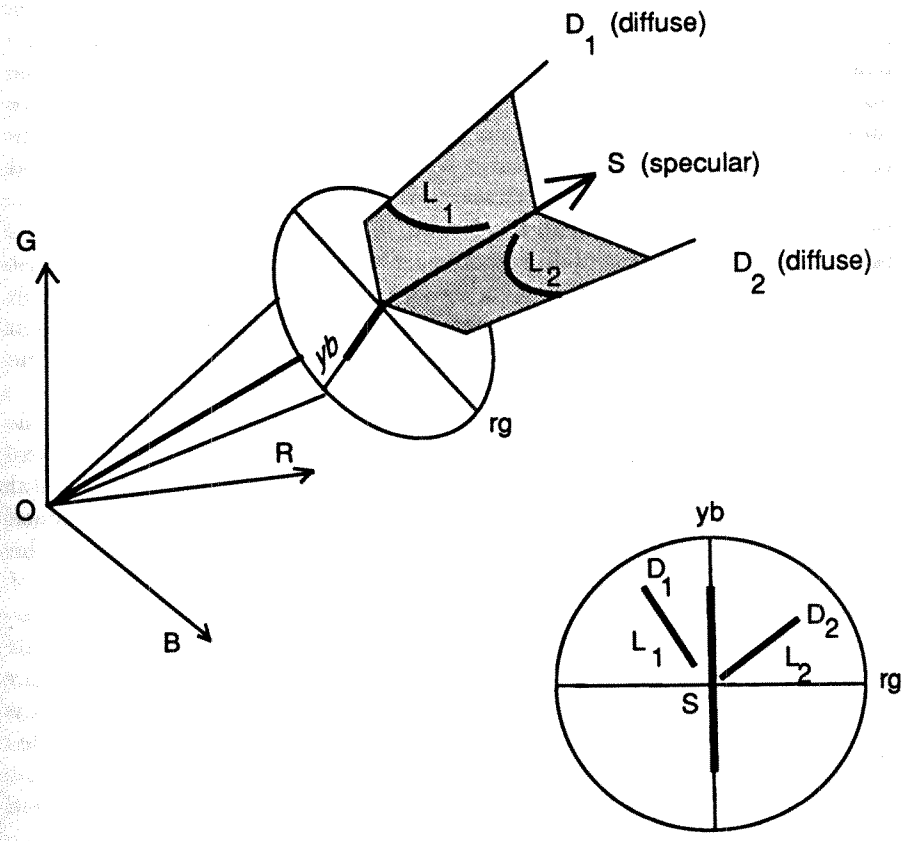


Fig. 13. Top: *Representation in the (R, G, B) space of responses L_1 and L_2 to two surface patches, lit by the same source S, that have different diffuse components of reflectance (D_1 and D_2). The two planes described by L_1 and L_2 intersect along the axis S, which describes the chromaticity of 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 and Lennie, 1986.)*

"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 and 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-monotonicity (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 behavior must be dealt with by the explicit use of contextual information (McDermott and 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 color 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, 1988). 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 perceiver). 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.

Acknowledgments

WR is supported by AFOSR 89-504 and AJ by NSERC Canada, IRIS Canada, and CIAR. We appreciate the helpful comments and issues raised by Jacob Feldman, Horace Barlow, Richard Mann and Donald Hoffman.