What Makes Viewpoint Invariant Properties Perceptually Salient?[1]

David W. Jacobs

Department of Computer Science

University of Maryland

djacobs@cs.umd.edu

It has been noted that many of the perceptually salient image properties identified by the Gestalt psychologists, such as collinearity, parallelism, and good continuation, are invariant to changes in viewpoint. However, we show that viewpoint invariance is not sufficient to distinguish these Gestalt properties; one can define an infinite number of viewpoint invariant properties that are not perceptually salient. We then show that generally, the perceptually salient viewpoint invariant properties are *minimal*, in the sense that they can be derived using less image information than non-salient properties. This provides support for the hypothesis that the biological relevance of an image property is determined both by the extent to which it provides information about the world and by the ease with which this property can be computed.

[1] Jacobs2000 contains an abbreviated version of this work, including technical details that we avoid in this paper.

1. Introduction

The Gestalt psychologists noted the privileged status of certain perceptual properties such as proximity, symmetry, good continuation, and similarity (eg.,[Koffka1935/1963]). These properties play a special role in early visual processes such as perceptual grouping, figure/ground discrimination, and pop-out. There have been many attempts to explain why these properties are especially salient. In one line of work, these properties are seen as key to producing the simplest organizations of images, perhaps in terms of efficient coding (eg.,[Attneave1954, Garner1962, Leeuwenberg1971]). In another line, these properties are seen as facilitating inferences about the scene that produced an image (eg.,[Witkin Tenenbaum1983, Lowe1985, Rock1983]). Specifically, some of these explanations of Gestalt phenomena focus on the non-accidentalness of these properties. A non-accidental image property is one that is unlikely to occur by accident, but which may occur due to some structure in the world. So it is a useful probabilistic clue about scene structure. These approaches are reviewed in,[Pomerantz Kubovy1986] and we draw upon that article in our own discussion.

In this paper we will speak of non-accidental properties (NAPs) as properties of sets of image features, and the scene features that produce them (this is defined more formally in Section 2). If an NAP is viewpoint invariant, it means that some sets of scene features always produce this property, regardless of the observer's viewpoint while all other sets of features produce this property only from chance viewpoints, if at all. In this case, the 2-D image property is deemed unlikely to occur by chance, but likely to occur because the 3-D scene contains a structure that always produces images

with this feature. For example, if a straight line appears in the image, this might be due to a straight line in the 3-D scene, which will always appear straight in the image. Or, it might be due to a planar curve that is viewed end-on. The unlikeliness of such an end-on view may license an inference in favor of the first possibility. In general, viewpoint invariant features provide a means of probabilistically inferring the 3-D structure of the scene from a single 2-D image.

Several authors (Witkin Tenenbaum1983, Binford1981, Kanade1981, Lowe1985) note that many of the most salient Gestalt grouping clues, such as connectivity, parallelism, symmetry, and good continuation, are instances of viewpoint invariant properties (some of these properties are illustrated in Figure 1). They then suggest that it is the leverage these 2-D image properties provide for inferring 3-D structure that explains their perceptual salience (Cutting1983, Van Gool etal1994 provide further discussion of the role of invariance in vision)., Lowe1985 and, Biederman1987 argue that viewpoint invariant properties play a key role in object recognition, providing a means of quickly matching 2-D images to known 3-D objects. The exact nature of the role of viewpoint invariance in object recognition is further discussed in, for example, Jolicoeur Kosslyn1983, Rock DiVita1987, Corballis1988, Biederman Gerhardstein1993, Tar and. Kurbat1994 There is strong evidence that performance on many recognition tasks, especially recognition of objects after a rotation in depth, are strongly effected by the presence or absence of viewpoint invariant properties (this issue is discussed in Hayward Tarr1997 and; Biederman Bar1999, Palmer1983 also reviews earlier evidence for this). Much work therefore connects the salience of viewpoint invariance to performance on the task of relating 2-D images to 3-D structure, suggesting one reason for this salience.

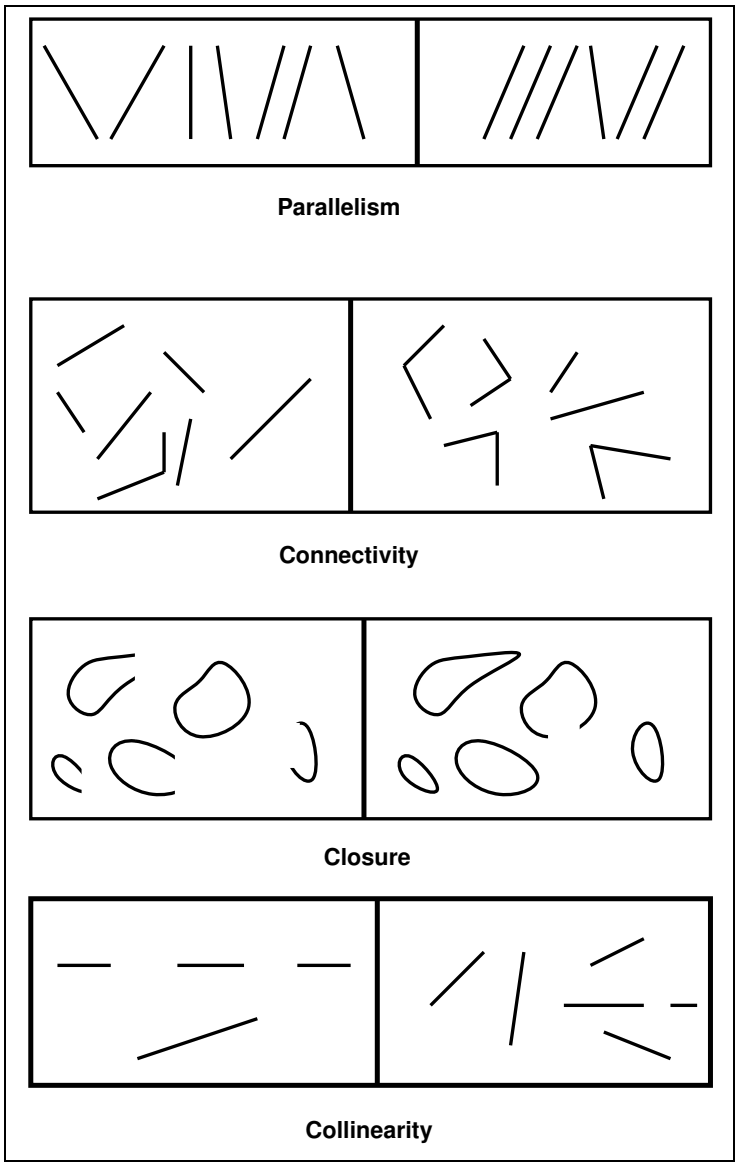

Fig. 1. Four Gestalt image properties are illustrated. From top to bottom: parallel lines, connected lines, collinear lines and closed curves are shown in contrast to other lines and curves that do not have these properties.

In this paper we examine viewpoint invariance as an explanation for perceptual salience. We show that there exist an infinite set of viewpoint invariant properties, almost all of which are not perceptually salient. We then show that salient viewpoint invariant properties do share an interesting quality. When the number of image features used to construct an image property is limited, only a small number of viewpoint invariant properties emerge. We call these *minimal* viewpoint invariant properties, and show that they are the ones that are perceptually salient.

Our approach naturally handles the Gestalt properties of closure, connectedness, parallelism, collinearity, convexity, and corners and trihedral vertices. These are all readily explicable as minimal viewpoint invariants. And we show that when viewpoint is restricted to be that of an upright observer, vertical and horizontal orientations are minimal viewpoint invariants. We can also explain symmetry as a viewpoint invariant property that is not minimal, but that is directly derived from a minimal property. This is the basis of the model of,[Wagemans etal1991] and,[Wagemans etal1993] which we draw on. Also, we consider other Gestalt properties, such as proximity, and relative smoothness, which can be seen as having a minimal viewpoint invariant property, such as that of being identical, as their limiting case. These properties are based on minimal viewpoint invariants, and provide inferential leverage, without being completely invariant.

The importance of minimality in perceptual salience provides evidence for the importance of computational, or process-based considerations in perceptual salience. Minimal properties are more efficiently detected than properties that rely on more image features. If complex properties, such as symmetry, can be derived as the con-

junction of simpler properties, they may also be more efficiently detected. Minimal properties can also be captured by simpler models of shape than are required to capture higher order properties, and these models can be more efficiently encoded and learned. Therefore, accepting the importance of minimality in perceptual salience leads to research directions based on the expectation that the computational complexity of building up and using image descriptions plays a key role in determining whether and how they will be used.

2. Preliminaries

We begin by defining six key terms: **image properties**, **non-accidental properties**, **inferential leverage**, **viewpoint invariant properties**, **perceptual salience**, and **minimal properties**. These fit together in our claim that viewpoint invariant properties that are perceptually salient are the ones that are most useful for accomplishing perceptual tasks. Their utility stems from the inferential leverage that their non-accidentalness provides, as has been previously noted. But in addition, perceptually salient properties almost always have a minimal number of features, which enhances their computational value.

Image Property We define an image property to be any binary function of an image, or part of an image. That is, given part of an image, we can say that that subimage either has, or does not have, that property. For example, parallelism is a property of pairs of lines.

Non-accidental properties Non-accidental properties are defined to be properties

that are likely to occur as a result of scene structure, but unlikely to occur due to a random background process. Therefore these properties can be used to infer the presence of scene structure. For example, suppose leaves float from trees down to the ground in a random way, but that occasionally a thin straight branch falls to the ground, with leaves still arranged about the branch in a regular pattern, such as along a straight line. Then, if one glances at the ground and spots a set of leaves in a row, one can infer that these fell together, along with a branch, even if one hasn't spotted the branch that connects them. This is because the image regularity of collinear leaves will occur due to a scene structure, but is very unlikely to occur as the result of independent leaves randomly falling to the ground.

More formally, we can say that a non-accidental property has zero probability of occurring due to a specific random process, but a non-zero probability of occurring due to scene structure.[2] In this case, it is certain (there is probability one) that the property occurred due to scene structure. Note, though, that a property is non-accidental only with respect to a well-defined random process and scene structure.

Inferential Leverage Therefore, non-accidental properties provide great inferential leverage. In general, we may say that a property has inferential leverage when its

[2]Note that a specific event may occur with zero probability even though its probability density is not zero. For example, if we choose a real number between zero and one from a uniform distribution, we could pick any real number in this interval, and the probability density is one over this interval, but any given number has zero probability of being chosen.

presence or absence allows us to make a probabilistic inference about the world. As we have defined non-accidental properties, they provide maximum inferential leverage, since they tell us something certain about the world. If a property's presence or absence tells us nothing that we didn't already know about the world, it provides no inferential leverage. In between these two extremes are clues that provide degrees of probabilistic information about the world, telling us that something in the scene is relatively more or less likely than we previously could suspect.

Viewpoint Invariant Properties We will focus on viewpoint invariant non-accidental properties. The non-accidentalness of these properties comes from the fact that some sets of 3-D scene features almost always produce these image properties while all other sets of scene features almost never produce these properties. That is, the non-accidentalness is derived directly from the fact that the presence or absence of this property is almost never due to viewpoint, but rather due to an underlying scene property. For example, parallelism will be a viewpoint invariant property for viewing transformations such as orthographic projection in which parallel 3-D scene lines always appear parallel in the image, while all non-parallel scene lines almost never appear parallel. Therefore, in our terminology, viewpoint invariants are always non-accidental properties. We will often implicitly exclude *trivial* viewpoint invariants. For example, the property of being an image is a viewpoint invariant (all images produced by scene features are images) but it is a trivial one, since it occurs in all cases.

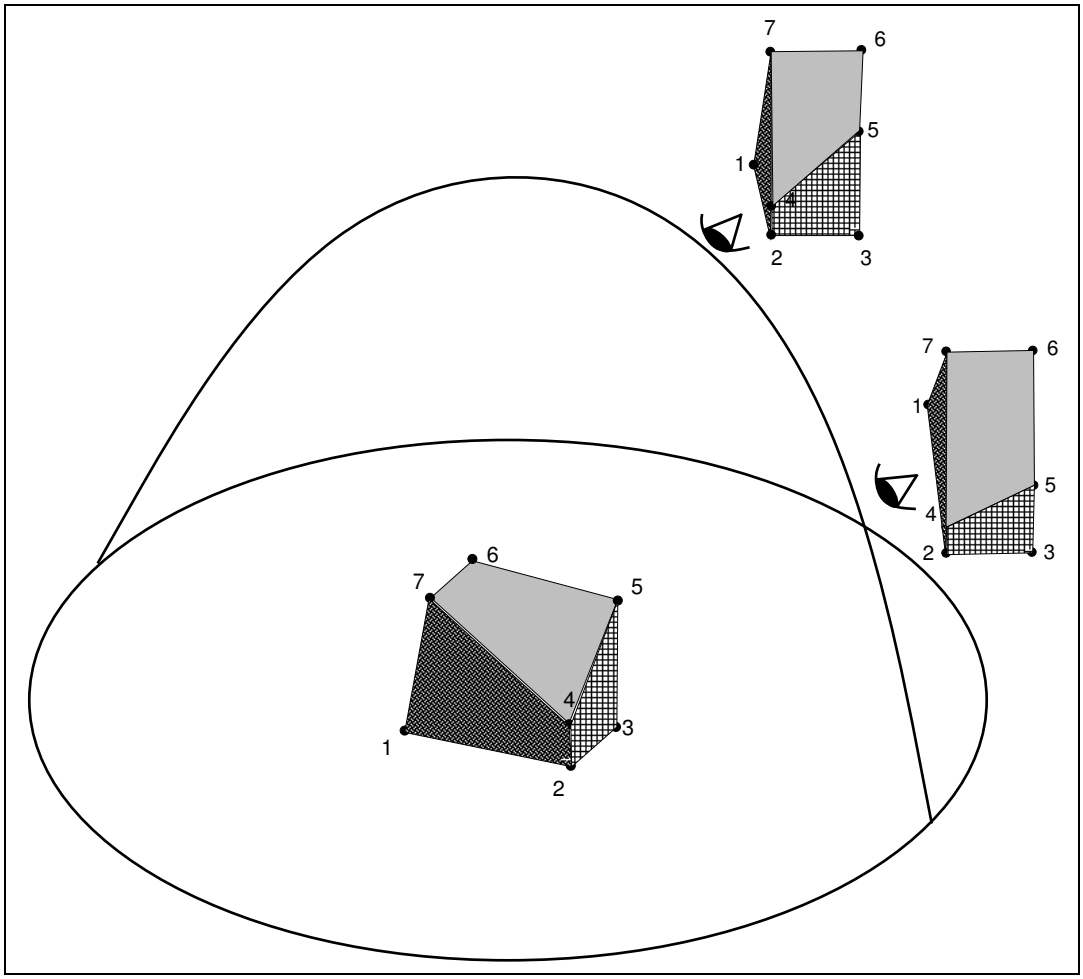

Fig. 2. A polyhedron is shown in the center. One can see that the two lines connecting points four and seven, and points five and six are not parallel in 3-D. However, there will be a circle of possible viewing directions from which these lines will appear parallel. Two such views are shown in the top of the figure. The lines will not appear parallel except from viewing directions lying on this circle. The figure also shows lines that are not collinear in 3-D, but that will appear collinear from these views.

To be more precise, by *almost always*, we mean with probability one, and by *almost never* we mean with probability zero. For example, if we throw two line segments into an image at random, there is zero probability that they will be perfectly collinear. We use almost always (never) to distinguish such unlikely events from ones that are impossible in principle.

To be formal about this, we must specify some probability distribution on the set of possible viewpoints. Then we can say that an object will produce a particular property with probability zero if and only if the probability of randomly selecting a viewpoint from which the object's image will have this property is zero. This is illustrated in Figure 2. It turns out that the specific choice of a probability distribution on viewpoints is not too important. Any two distributions will give rise to the same viewpoint invariants as long as one of the distributions does not make a specific set of viewpoints infinitely more likely than the other distribution does. [3]

[3]To avoid confusion, the reader should note that the term invariance is used with somewhat different meanings by different authors, and common usage varies within the fields of mathematics, perceptual psychology, and computer vision. In mathematics, there are some further technical restrictions on the meaning of an invariant. Most importantly, invariants are considered relative only to possible transformations that form a *group*. For transformations to form a group we must be able to apply two transformations, one after the other, and have this be equivalent to applying a third transformation in the group. Also, the inverse of each transformation must also be in the group. Sometimes the group structure of the transformations is also a requirement of invariants considered in perceptual psychology and computer vision (eg.,[Palmer1983, Weiss1993]). However, this is only possible when invariants are considered relative to a transformation that starts with a 2-D object and

Perceptual Salience We consider image properties perceptually salient when they play a special role in early visual processes. This is clearly quite a broad definition that can be manifested in many different ways. In particular, we generally mean that the contrast between the presence or absence of this property is what can be salient. These contrasts are the basis of Biederman's recognition-by-

produces a 2-D object. Viewing transformations that turn the 3-D world into a 2-D image do not form a group because they are not invertible and cannot be composed. For this reason, we adopt a definition of invariants that does not require transformations to form a group.

Also, *invariant* is often used to mean that an image property never changes under projection, while we use it to mean that the property almost never changes. One reason for this is that it has been shown ([Burns etal1992, Clemens Jacobs1991, Moses Ullman1992]), that general 3-D objects do not have completely invariant properties when viewed in a single 2-D image. Properties that are completely invariant do exist when the set of possible objects is limited, as when objects are assumed to be planar and never viewed end-on (as described in Section A) or objects are rotationally symmetric (eg.,.[Forsyth etal1992] When the universe of possible objects includes all objects, or even a reasonably wide range of objects, then there can be no completely invariant image properties. The technical details of these limitations are discussed at more length in[Jacobs1997] and in.[Basri Moses1999] Jacobs shows, for example that a set of objects each consisting of five 3-D points can produce an invariant if and only if the heights of the fourth and fifth points above the plane of the first three have the same ratio for all point sets. Jacobs also shows that viewpoint invariants can be constructed in general only for sets of objects that are a measure zero subset of the set of all possible objects. Basri and Moses show that invariants are possible only when a subset of objects is considered so that the object can be reconstructed from a single image. In considering properties that are almost always preserved under projection from 3-D to 2-D we are focusing on the properties that we feel are most natural in vision, while reminding the reader that many researchers have considered different definitions of invariants from different perspectives.

component theory of the use of non-accidental-properties in recognition (eg.,[Biederman1987]). Or, in referring to good continuation as a perceptually salient property, we mean that either continuity or discontinuity may play a role in early visual processes such as grouping or pop-out. It is also true that whenever the presence of a property is viewpoint invariant, the absence of that property will also be viewpoint invariant. So, for example, it follows directly that if parallelism is always produced by parallel lines and almost never produced by non-parallelism, then the absence of parallelism is always produced by non-parallel lines, and never produced by parallel ones.

One may also talk about the degree of perceptual salience possessed by an image property. For example,,[Van der Helm Leeuwenberg1996] discuss the relative salience of various forms of symmetry, such as mirror symmetry, skew symmetry, and the symmetry of glass and repetition patterns. And proximity and good continuation can be thought of as properties that are possessed to varying degrees, with greater salience when, for example, features are closer together or curves are more smoothly continued (eg.,[Field etal1993]). While we do not present an extensive treatment of degrees of salience, we will discuss this issue in Section C. In these cases, it may be more appropriate to speak of the presence of a property, such as proximity, as being salient, while its absence is not particularly salient. In referring to Gestalt properties as salient, then, what we really mean is that the detection of these properties, or their absence, is the building block for low-level perceptual performance. The visual system seems especially sensitive to

these properties, and especially influenced by their presence or absence.

Minimal Properties One of the main contentions of this paper is that the viewpoint invariant properties that are perceptually salient are the ones that are based on a minimal number of features. This is probably best explained through the examples that will follow. We will show that for simple features such as points or lines, once we consider properties of a large number of features, an infinite number of these properties will be viewpoint invariant. However, for a small number of features, there will only be a small number of properties that are viewpoint invariant. We will refer to properties based on such small sets of features as minimal properties.

3. Viewpoint Invariant Properties

We now describe the set of viewpoint invariant properties that can be derived from scenes and images. We will begin with point features, which are the easiest features to understand mathematically. We will consider images and scenes composed only of points in some detail, not just because of their intrinsic interest, but also to illustrate how we can analyze viewpoint invariants. We will show how an infinite set of viewpoint invariant properties exist for points, and then describe the minimal viewpoint invariant properties. Next we will use these results as a basis for analyzing properties of curves and lines. Finally, we will discuss the relevance of these results in understanding the geometric features produced in images by 3-D shapes.

A. Viewpoint Invariant Properties of Points

In this section we will suppose that a scene is composed of 3-D points, and that an image consists of the 2-D projection of these points. This is the idealization of the world that is simplest to thoroughly understand, and it also captures much of the essence of scenes consisting of polyhedral objects, which are well described by the location of their vertices. We let $\mathbf{p} = (x, y, z)$ describe a typical scene point, with $\mathbf{q} = (u, v)$ a typical image point. We will consider two types of projection as models of the image formation process. The first is a standard pinhole model of perspective projection. The second projection model, called *scaled-orthographic projection*, is an approximation to perspective projection. One way to think of the difference between these two projections is that in true perspective, the (x, y) coordinates of each point are scaled inversely according to their distance to the image plane. In scaled orthogaphic projection, all points are scaled by the same factor, which can depend on their average distance to the image. This will be an accurate approximation when the distance to all points is similar relative to their distance to the camera.

Although we consider scenes composed of sets of 3-D points, it will be important to first note that in the special case when the scene points are coplanar, the set of projections that act on them have a simpler form. Specifically, the scaled orthographic projections of a set of planar points are equivalent to the 2-D affine transformations. These are just linear transformations of the points, along with possible translation of them relative to the origin. The perspective projections of a planar point set can be modelled by a subset of the projective transformations. Projective transformations are

the group of transformations formed by allowing perspective projection with points either in front of or in back of the image plane, and by allowing all compositions of these projections. Elegant mathematics is available to analyze projective transformations. A full description of these models for planar projection can be found in standard geometry texts, such as Tuller, 1967. In this paper we will simply state and make use of a few basic properties of these projections, while avoiding mathematical derivations.

This paper relies on one key mathematical fact. A set of planar points can be described using a representation that is invariant under 2-D affine transformations, or that is invariant under projective transformations. In the affine case, we can use the first three non-collinear points to define a new coordinate system. Specifically, we let the first point, $\mathbf{p_1}$ be the origin of this new coordinate system. The direction from $\mathbf{p_1}$ to $\mathbf{p_2}$ describes one axis of the coordinate system, which we'll call u, while the direction from $\mathbf{p_1}$ to $\mathbf{p_3}$ will determine the other axis, v (we use u and v to distinguish these from x and y, the traditional axes of a Cartesian coordinate system). Lengths in each direction are defined so that the distance from $\mathbf{p_1}$ to $\mathbf{p_2}$ or $\mathbf{p_3}$ is 1. So $\mathbf{p_2}$ has coordinates $(1, 0)$ in this coordinate system, and $\mathbf{p_3}$ has coordinates $(0, 1)$. When we use the points $\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}$ to define a coordinate system in this way, they may be referred to as a *basis*.

Then we may describe any additional point by its coordinates in this system, which we call *affine coordinates*, for reasons we will give shortly. For example, let (α_4, β_4) denote the affine coordinates of the fourth, coplanar scene point. This means

that:

$$\mathbf{p_4} = \mathbf{p_1} + \alpha_4(\mathbf{p_2} - \mathbf{p_1}) + \beta_4(\mathbf{p_3} - \mathbf{p_1})$$

This is in direct analogy to Cartesian coordinates, for if $\mathbf{p_4}$ has cartesian coordinates (x, y) then:

$$\mathbf{p_4} = (0, 0) + x(1, 0) + y(0, 1)$$

The affine coordinates of the fourth scene point are unaltered by any 2-D affine transformation. Therefore, when this scene is projected to form an image, the fourth image point will have the same affine coordinates relative to the first three image points that the fourth scene point has relative to the first three scene points. That is, recalling that $\mathbf{q_i}$ is the image of $\mathbf{p_i}$, we always have:

$$\mathbf{q_4} = \mathbf{q_1} + \alpha_4(\mathbf{q_2} - \mathbf{q_1}) + \beta_4(\mathbf{q_3} - \mathbf{q_1})$$

In a similar fashion, we may describe a fifth, coplanar scene point's position relative to the first four in a way that is invariant under perspective projection. These results are well known, and the reader may find a fuller explanation in.[Tuller1967]

1. Viewpoint invariance alone does not explain perceptual salience

We now show that an infinite number of viewpoint invariant properties exist, most of which are not perceptually salient. We begin by considering the case of a scene and image containing four points, under scaled-orthographic projection. We will define an infinite class of properties for this setting, and show that all of these properties are viewpoint invariant. This will rely on the fact that for planar point sets, an affine invariant property is preserved under (almost all) viewpoints. As we will show, this

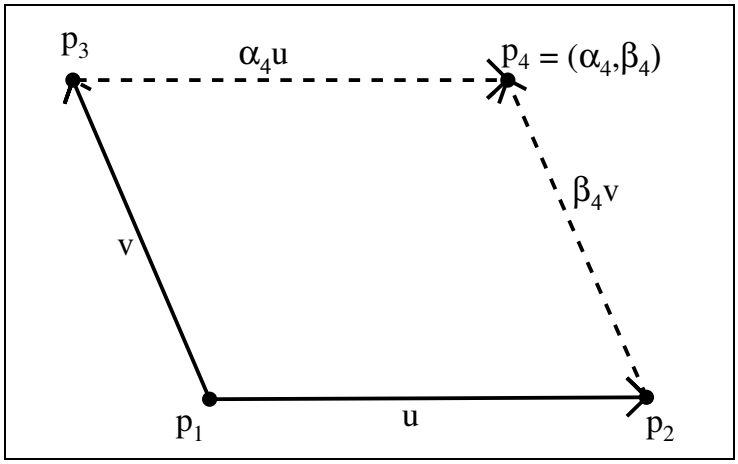

Fig. 3. $\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}$ form an affine basis. $\mathbf{p_4}$ has coordinates (α_4, β_4) with respect to this coordinate system. The points form a parallelogram, for example, when $(\alpha_4, \beta_4) = (1, 1)$.

implies that affine-invariant properties are also viewpoint invariants, as long as any given non-planar point set rarely produces this property. The key point that we show here is that properties that would be strictly invariant when we limit our universe of possible objects to planar ones become non-accidental properties when we consider more general sets of 3-D objects.

We will say that four image points have the property P_{α_4,β_4} if it is possible to order them so that the fourth point has the affine coordinates (α_4, β_4) relative to the first three. To provide a better intuition of these properties, we note that four image points have the property $P_{1,1}$ if and only if they form a parallelogram (see Figure 3).

We can show that all properties of the form P_{α_4,β_4} are viewpoint invariant. One further mathematical fact is first needed. Choose a specific ordering of a set of four non-coplanar scene points, and a specific pair of affine coordinates. Then there exist exactly two viewpoints from which these scene points produce an image having this

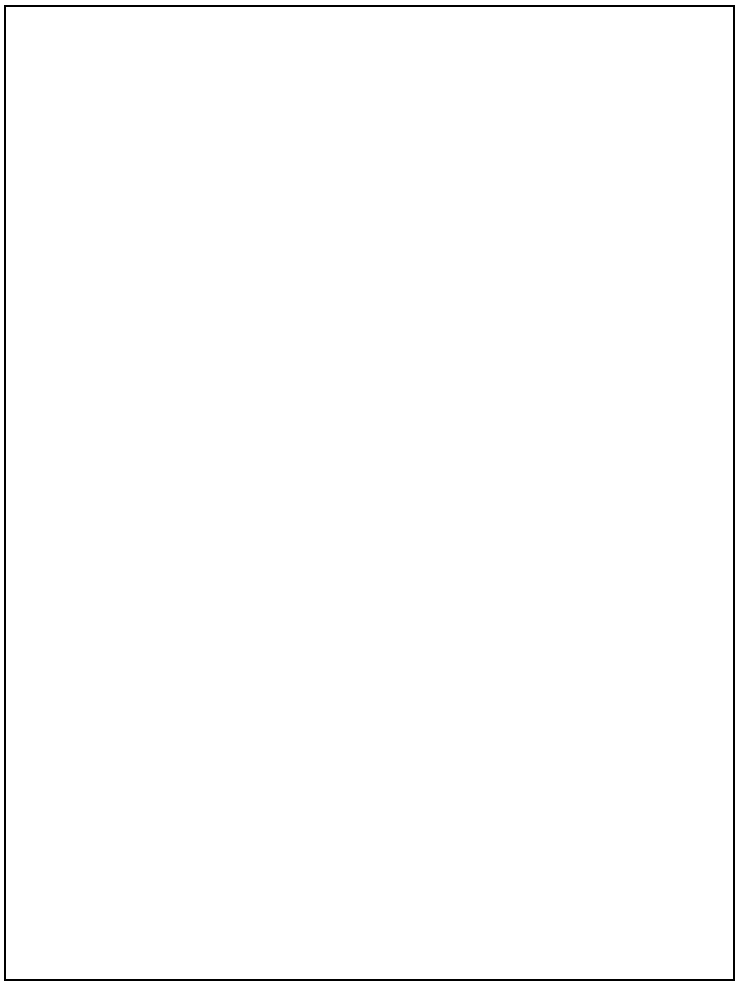

Fig. 4. We define the scene plane as the plane formed by p_1, p_2, p_3. For any affine coordinates, (α_4, β_4), some point in this plane, s_4, has these coordinates. The fourth scene point, p_4, projects to the same image point as s_4 from one viewing direction, as shown. So, from this direction, p_4 appears in the image with affine coordinates (α_4, β_4).

pair of affine coordinates relative to the first three image points. (These viewpoints differ by 180 degrees, so that the points seen from one view are just the reflection of the points seen from the opposite view). This was shown in.[Clemens Jacobs1991] Briefly, as shown in Figure 4, we first define the *scene plane* as the plane formed by the first three scene points. Some point, $\mathbf{s_4}$, in the scene plane has affine coordinates (α_4, β_4) relative to the first three scene points. If we form a line including $\mathbf{s_4}$ and $\mathbf{p_4}$ (the fourth scene point), this line describes a viewing direction from which $\mathbf{p_4}$ and $\mathbf{s_4}$ project to the same image point, $\mathbf{q_4}$. Since $\mathbf{s_4}$ is coplanar with the first three scene points, it has the same affine coordinates when viewed from any direction, so $\mathbf{q_4}$ has affine coordinates (α_4, β_4). As a simple example, suppose our points have coordinates: $(0,0,0), (1,0,0), (0,1,0), (1,1,1)$. Notice that the first three points lie in the $z = 0$ plane, and that the point $(1,1,0)$, also in this plane, has affine coordinates $(1,1)$ relative to the first three points (in this simple example, the affine and Cartesian coordinates are identical). When we view all four points directly from above or below, the fourth point, $(1,1,1)$, has the same position in the image as $(1,1,0)$, and so it also has affine coordinates $(1,1)$. When viewed from any other position, $(1,1,1)$ appears at a point in the image that is different from the position of $(1,1,0)$, and so has affine coordinates that are not $(1,1)$.

Therefore, for any property P_{α_4,β_4}, every set of four non-coplanar points will produce an image with this property from only two viewpoints, ie. with probability zero. A set of planar scene points will almost always produce this property if the fourth point has affine coordinates (α_4, β_4). If four coplanar points are looked at from a viewpoint lying in the plane formed by the points they will appear collinear. From

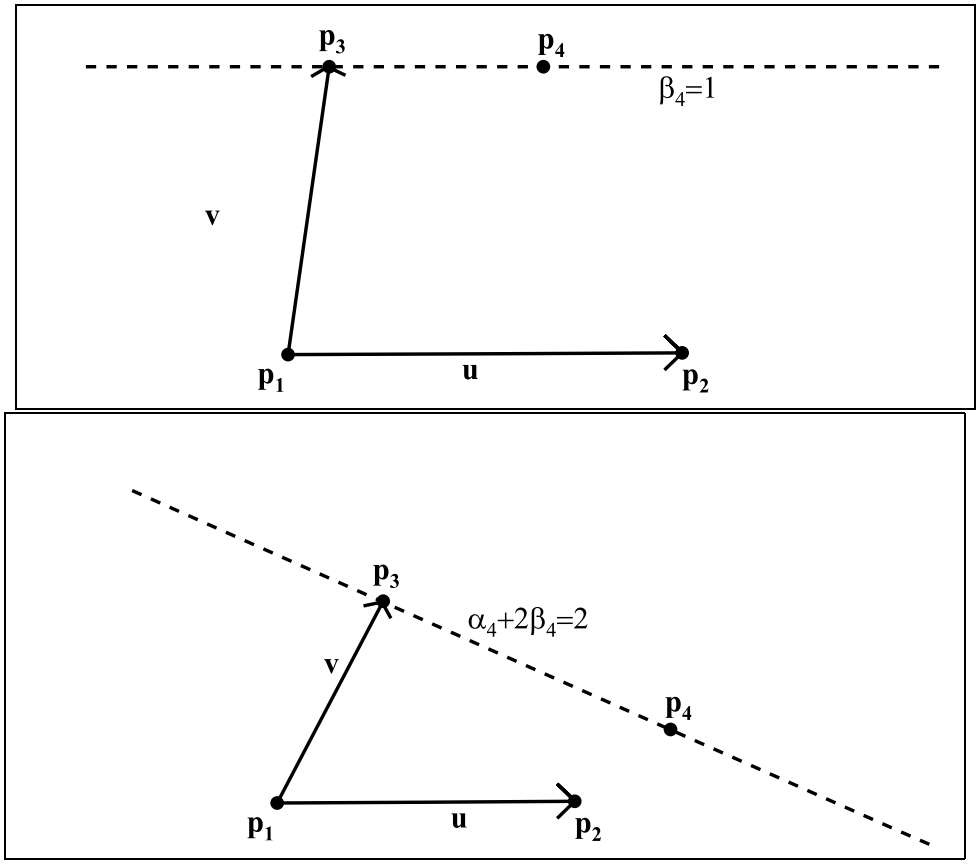

Fig. 5. $\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}$ form an affine basis. On the top, the dashed line is described by $\beta_4 = 1$. On the bottom, the dashed line is described by $\alpha_4 + 2\beta_4 = 2$. If $\mathbf{p_4}$ falls on the line on the top, the four points form the corners of a trapezoid. On the bottom, a point on the dashed line has an equally viewpoint invariant property which is not, however, perceptually salient.

this set of viewpoints their affine coordinates will be undefined, but these viewpoints have probability zero of occurring. Otherwise, the points will produce a fourth image point with affine coordinates (α_4, β_4). On the other hand, if four coplanar points do not have affine coordinates (α_4, β_4), they will never produce an image with these affine coordinates. Possible scenes are therefore divided into two classes, those that almost always produce the property, and those that never, or almost never, do.

We may extend these results to produce a wider class of viewpoint invariant properties, as well. We will consider one example in detail. First, we note that the α-β coordinates give us a simple, abstract description of a configuration of four points in the plane. We may represent a point-quadruple (ie., a configuration of four points) as just a single point in a 2-D α-β space. This point (a pair of affine coordinates) tells us everything essential about this point-quadruple because everything else about the position of the four points is completely variable with viewpoint. Only the α-β coordinates capture aspects of the points that are viewpoint invariant.

,[?] in an influential work proposed the use of trapezoids (ie., quadrangles with two parallel sides) in a computer perceptual organization system, to illustrate the value of viewpoint invariant non-accidental properties. The property of being a trapezoid is a viewpoint invariant property because two parallel scene lines always project to parallel image lines, under scaled-orthographic projection, while non-parallel scene lines appear parallel only from special viewpoints. Without loss of generality, pick three of the corners of a trapezoid to describe an affine basis, with the α coordinate given by one of the two parallel lines. Then a set of four planar points is a trapezoid if and only if $\beta_4 = 1$ (see Figure 5). $\beta_4 = 1$ implies that the line from $\mathbf{p_3}$ to $\mathbf{p_4}$ will be parallel to the line from $\mathbf{p_1}$ to $\mathbf{p_2}$. This is because $\mathbf{p_4}$ and $\mathbf{p_3}$ have the same β value, and so the difference between them is only in the direction of the α axis, and the same is also true of the points $\mathbf{p_2}$ and $\mathbf{p_1}$. So, abstractly, in the space of possible (α_4, β_4) coordinates, the set of trapezoids is defined by the line $\beta_4 = 1$. We may call this the property $P_{\alpha,1}$, where α_4 takes on any value, but β_4 must be one.

This description of trapezoids as a line in α-β space makes clear that they are a

non-accidental property. Any set of four non-coplanar 3-D scene points will produce images with all possible values for α_4 and β_4. And only a one-dimensional subset of these values represent images that are trapezoids. These come from only a special subset of viewpoints which will occur with zero probability. On the other hand, if four points are coplanar and form a trapezoid, their affine coordinates will be invariant, with β_4 always equal to one, except for degenerate views.

However, we may now see that any line in (α_4, β_4) space also defines a viewpoint invariant property. Figure 5 shows images with the property defined by $\alpha_4 + 2\beta_4 = 2$, for example. Planar point sets with affine coordinates on this line almost always produce images with affine coordinates on this line, since their affine coordinates are viewpoint invariant except when the points are viewed end-on. Non-planar point sets will produce several images[4] with the affine coordinates of any point on this line, along with several images with any other possible pair of affine coordinates. Again, the images they produce with the viewpoint invariant property described by such a line is a 1-D subset out of a 2-D set of possible images. Therefore, a 3-D set of points will produce affine coordinates that lie on the line $\alpha_4 + 2\beta_4 = 2$ only with probability zero. So the property of four image points having affine coordinates that satisfy the equation $\alpha_4 + 2\beta_4 = 2$ is a viewpoint invariant. All in all, we can see that the vertices of trapezoids are not distinguished by viewpoint invariance per se, in comparison to the vertices of any other simple class of quadrangles. There are an

[4]More precisely, since there are 4! orderings of the points and two opposite viewpoints for each ordering that produce a particular set of affine coordinates, there will be 48 such images, unless symmetries cause some of these to be duplicates.

infinite number of properties with exactly the same degree of viewpoint invariance. The perceptual salience of trapezoids cannot therefore be explained solely by their viewpoint invariance, compared to other possible properties of quadrangles. As we will see, trapezoids can be distinguished from other viewpoint invariant characterization of quadrangles because they possess parallelism, a minimal viewpoint invariant.

Similar reasoning can be applied for the case of scenes containing more than four points undergoing either scaled-orthographic projection or perspective projection. These cases are discussed in some detail in.[Jacobs1997, Jacobs1996b]

In general, we can see that there is no small, distinguished set of viewpoint invariant properties based on four points. There are an infinite set of such properties which are on an equal footing as viewpoint invariants. This implies that viewpoint invariance is not a sufficient condition of perceptual salience.

2. "Minimal" viewpoint invariant properties are perceptually salient

Continuing with the simplest case of point features, we now show that there are some viewpoint invariant properties of point features that are mathematically distinctive. These properties are based on sets with only two or three points. We will show that for these small sets of points, most configurations do not have viewpoint invariant properties. But there are special points sets that have a mathematical degeneracy making them viewpoint invariant, and these point sets correspond to the properties of contiguity and collinearity, which are perceptually salient properties. Moreover, we will show that when three points lie on the boundary of a surface, the convexity or concavity of the points is also a viewpoint invariant property. We will call these

properties minimal, because they are based on small sets of features. In fact, our definition of a minimal property will be a property based on a set of features that is small enough so that general sets of these features do not have viewpoint invariant properties. We will defer until later discussion of our hypotheses about why these minimal viewpoint invariant properties would have a special perceptual status.

First, consider the case of scenes consisting of just two points. We begin by showing that the contrast between identity and difference is a viewpoint invariant property in this case. We say that a pair of image points display the property of identity if they appear in the same image location. If two scene points are identical, then it is obvious that they will always appear in the same position in the image. Suppose, on the other hand, that two scene points are distinct. In that case, they will appear at the same location in the image if and only if they are viewed so that they line up exactly in the image. That is, the two points define a line, and they must be viewed along this line to appear at the same image point. These special viewpoints will appear only with probability zero. Therefore, identity is a viewpoint invariant property of pairs of points.

As for all viewpoint invariant properties, if the presence of identity is a viewpoint invariant, so is its absence. Our definition of viewpoint invariance is symmetric in this regard. If some scenes almost always produce a property, then they almost never produce the absence of this property. And if some scenes almost never produce a property, they almost always produce the absence of this property. Therefore, it is the contrast between identity and difference that is a viewpoint invariant, although for the sake of brevity we will sometimes refer to this contrast by referring to one half

of it.

Next, it is possible to show that identity (in contrast to difference) is the only viewpoint invariant property for pairs of points. We will avoid the technical details that would be necessary for a rigorous proof (see[Jacobs2000] for these details), but instead provide an intuitive argument as to why this is true. Any two non-identical scene points can produce any possible pair of image points[5]. Moreover, for reasonable choices of probability distributions on the possible viewpoints, the probability of any particular image of the points will be not too different from the probability of any other image occurring. Given this, if there is some image property that a particular configuration of two scene points will produce with probability one, it must be a property that is true of almost all images. From this, it follows that any other configuration of distinct scene points will also produce images that have this property with probability one. So, if there is an image property that one scene produces with probability one, then all scenes (with two distinct points) will produce it with probability one. This means that it is a trivial, uninteresting property, such as just the property of being an image. This leaves identity as the only non-trivial viewpoint invariant property of pairs of points.

The property of two points being identical may not seem very meaningful, since

[5]This statement is only strictly true using the scaled orthographic approximation to perspective projection, in which scene points may be scaled arbitrarily in the image. With perspective projection, points may not project to be larger in the image than they are in the scene. In brief, taking account of this would not give rise to any new viewpoint invariant properties because while tiny objects can never appear too big in the image, large objects can always appear small.

two scene points that project to the same image point will be visible as only a single point. However, this property is basic to the properties of continuity and connectedness. For example, two line segments are connected if and only if their endpoints are identical.

Moreover, we can model closure as being derived from connectedness, and show that it is also a minimal viewpoint invariant property. To do this, we will temporarily switch from considering point features to considering contours. We distinguish between open contours, which have two endpoints, and closed contours. Then the fact that the two endpoints of an open contour are not identical is a minimal viewpoint invariant property. There will be a pair of viewpoints from which the open contour's endpoints appear in the same position in the image, and the contour appears closed (if the contour is planar then even this is not possible, since from this viewpoint the contour will look like a line segment). Of course, a closed contour always appears closed. Therefore, the viewpoint invariance of closure is derived from the minimal viewpoint invariance of identity.

This may seem odd, since closure is often described as a global property of contours, not a property derived from pairs of points (eg.,[Julesz1993]). Closure may seem global since one cannot say that a part of a contour is closed, in the same way that one says part of a contour is smooth. However, closure can be determined by the simple combination of local properties. A contour is closed if and only if no small portion of the contour contains an end point. Therefore, to determine closure, information must be propagated across the entire contour, but this information can be based on very local properties. In this way, closure is based on connectedness, a minimal viewpoint

invariant.

The same sort of argument can be used to show that when we consider configurations of three points, the only new viewpoint invariant property that emerges is collinearity. We first note that if three scene points are collinear, they will always project to collinear image points. On the other hand, suppose we have three non-collinear scene points. These points lie in some plane (since three points define a unique plane). They will appear collinear if and only if the viewpoint also lies in this plane. However, there is probability zero that a randomly chosen viewpoint will lie in a particular plane of the 3-D world. So collinearity, too, is a viewpoint invariant property.

On the other hand, we can also show that there are no other viewpoint invariant properties arising from configurations of three points. This is also because three non-collinear scene points can produce any configuration of three image points. Moreover, as was the case with identity, the probability that three scene points will produce any particular configuration of image points will not be that different from the probability of any other configuration arising, and so similar reasoning will show that no properties of three points, other than collinearity and identity, are non-accidental. It is possible to have a percept of smoothness or good continuation from triples of points (see eg.,[Feldman1997]). While this does not arise from a viewpoint invariant property of triples of points, we will discuss the related issue of smoothness of curves in Section B. We will discuss this issue overall more in Section C.

We must include the caveat that in addition to collinearity, there are viewpoint invariant characterizations of three points that are derived from identity. In the case

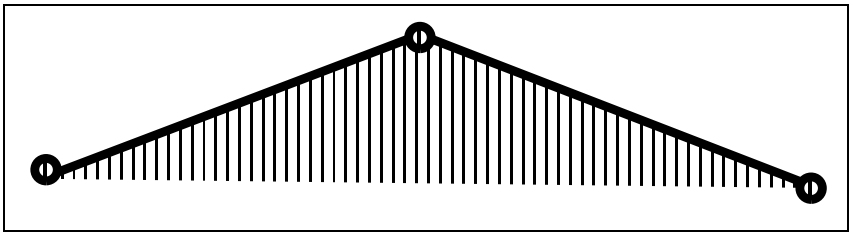

Fig. 6. Three points can be convex when one side of the curve joining them is distinguished as figure, and the other is background.

of three points, it is possible for any two of the points to be identical in the image, or for all three points to be identical. That is, the number of points is a viewpoint invariant description of the image.

Now, consider a set of three scene points that lie on a curve that forms the boundary of a surface. This gives us additional information: when we go from one point to the next we know which side of the curve the surface is on, and which side is background. For this discussion, suppose for example that the three points are consecutive vertices of a polygon. While the whole polygon need not be planar, three consecutive vertices will be coplanar since any three points that are not collinear lie in a unique plane. As we proceed from $\mathbf{p_1}$ to $\mathbf{p_2}$ to $\mathbf{p_3}$, the surface boundary will be either convex or concave, supposing that the three points are not collinear (Figure 6). This convexity or concavity will be a viewpoint invariant property of the points. This follows from the fact that if the points are convex, they will always appear convex, while if they are concave, they will always appear concave. This property is discussed at length by,[Hoffman Richards1984] who apply it to the problem of representing the shape of curves. We can also show that convexity/concavity is the only viewpoint invariant

property introduced by assuming that three points lie on the boundary of an object. In brief, one can show that three convex points can project to any convex positions in the image, while three concave scene points can produce any three concave image points, and then apply reasoning similar to that given above. Knowing figure and background was important in developing this viewpoint invariant property, because without this the concept of convexity is not well-defined for three points. For example, if the three points are vertices of a polyhedron that are on its interior, ie., they do not lie on the silhouette, then one cannot define them to be convex or concave in the image.

In summary, we can see that there exist an infinite number of viewpoint invariant properties for scenes with four or more points. But the situation changes dramatically for scenes with only two or three points. For these scenes, the only viewpoint invariant properties are collinearity, convexity, and identity. We refer to these as minimal viewpoint invariants, because they involve the smallest possible number of features. Since convexity, collinearity, and identity correspond to perceptually salient features, it is suggestive that there is some connection between minimality and perceptual salience.

We will discuss possible reasons for this connection in Section 4, after considering the viewpoint invariant properties to which other features give rise.

B. Differential properties of curves

We will now consider a more complex case, that of differential viewpoint invariant properties of curves that are the boundaries of objects. For example, the simplest differential properties of a curve are its tangent and curvature. Mathematically, the

differential properties of a curve at a point can be thought of as the properties of a series of equally spaced points near that point, in the limit as the distance between these points goes to zero. Therefore we can reason about differential properties in much the same way that we reasoned about properties of sets of points. In perceptual systems, due to discretization and noise in an image, these properties must be computed at a particular scale, and need not rely on equally spaced samples. We will ignore these effects in this paper, but consider only the ideal case of continuous, error-free curves.

First, we consider the first derivative of a curve. This is defined by the direction between two points, as the distance between them goes to zero (ie, the tangent). In the case of points, we reasoned that two distinct points could never possess a viewpoint invariant property because they could produce any possible image; similarly we can reason that two points that, in the limit, define a tangent, can never possess a viewpoint invariant property because any scene tangent can produce any image tangent, through a simple rotation. The only difference is that the degenerate case of two identical image points does not exist, since the tangent is defined by distinct points only.

Minimal viewpoint invariant properties do arise when we consider differential properties based on three points, ie., curvature. Three point features have one viewpoint invariant property: collinearity. There is a corresponding viewpoint invariant property based on curvature. When three points on a curve are, in the limit, collinear, this means that the contour has zero curvature at that point. This occurs in straight lines, or at the inflexion points of curves. This is viewpoint invariant for the same

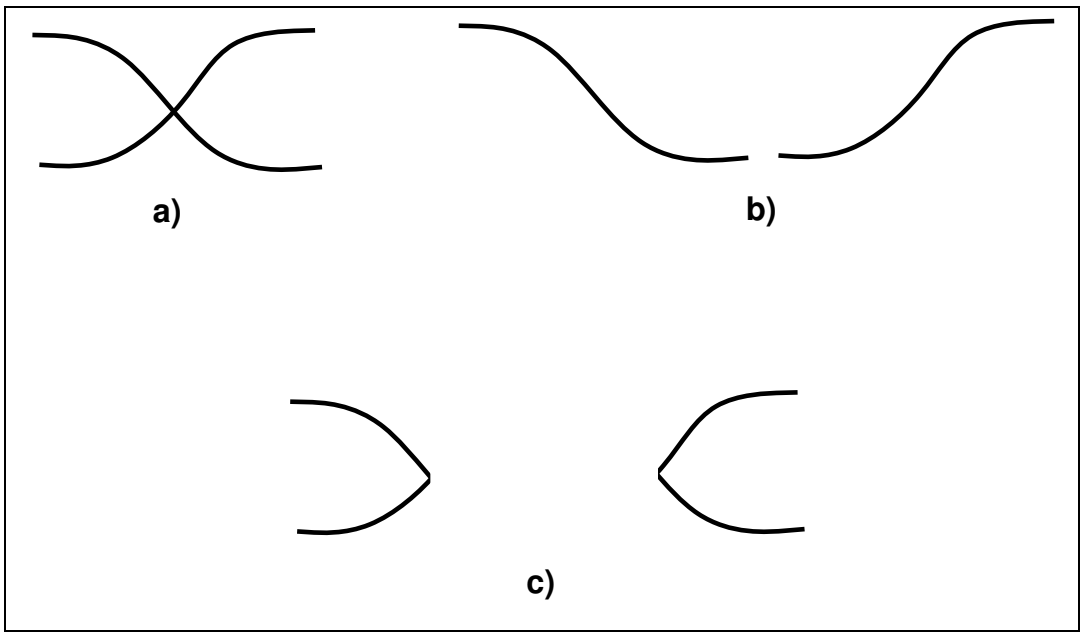

Fig. 7. As shown by Gestaltists, good continuation serves as a clue that the curves in a) are two smooth curves, as shown in b), not two curves with discontinuities, as shown in c).

reason that collinearity is viewpoint invariant. A contour point with zero curvature will always appear to have zero curvature, since the collinearity of three points is preserved under projection. A point with non-zero curvature can produce an image point with zero curvature only when viewed from within the plane formed by three points on the curve in the limit, as the distance between the points goes to zero. The situation is exactly analogous to that of triples of point features.

Also by analogy to the case of point features, we can show that when a curve has sides distinguished as figure and background, the local convexity or concavity of the curve will also be a viewpoint invariant property.

A new viewpoint invariant property arises, because the curvature can be infinite. This occurs when the contour point has a discontinuity in tangents. This property is

viewpoint invariant since a contour discontinuity almost always produces an image contour with a discontinuity, except when the contour is viewed from within the plane that the contour lies in locally. For those special viewpoints, the contour's image will have zero curvature. On the other hand, a curve that does not have a discontinuity will never project to an image containing a discontinuity. No other viewpoint invariants based on curvature exist, since a contour with positive non-infinite curvature can produce images with any positive curvature, through a simple scaling. Discontinuities in curvature are of interest not only because their presence signals a corner, but also because their absence indicates smooth curvature, which is one characteristic of good continuation, as shown in Figure 7. Good continuation can also be interpreted to mean a preference for less curvature, which we will discuss in Section C.

We can therefore divide the possible curvatures into three groups, zero, infinite and finite. Non-zero curvatures can be convex or concave, when one side of the curve is distinguished as figure, and the other as background. This gives us a full description of viewpoint invariant properties based on curvature.

While higher order differential properties of curves may be difficult to find reliably, it is worth noting that in principal we obtain a comparable situation to that found for points when we consider higher derivatives. The next higher-order property of curves is change in curvature, which reflects the behavior of four consecutive points on the curve. With this extra information, we may find differential properties of curves that are invariant to affine ([Cyganski etal1987, Weiss1993]) or projective ([Weiss1993]) transformations. As we did with points, we may then construct an infinite set of viewpoint invariant properties, which again, will not in general be perceptually salient.

In sum, differential properties of curves lead to a similar situation to that which arises with point features. A small set of viewpoint invariants exist when these properties must be based on a small amount of image information, either a few points or a few derivatives of a curve. These properties, based on minimal information, are generally perceptually salient. When we are allowed to use more points, or more derivatives, or both, we can obtain an infinite set of viewpoint invariants, which are not perceptually salient.

C. Lines

Similar reasoning can also be applied to the case of lines. As mathematical objects, lines are infinite in extent, but their presence is indicated in images by finite objects, such as line segments. It is worthwhile to analyze properties of lines when line segments in an image indicate orienation accurately, but do not have distinct end points (perhaps because of occlusion). Line segments with clear end points can be handled by our discussion of points, as applied to their end points. We can show for lines, as we did for points, that given two lines, the only viewpoint invariants that exist are identity and parallelism. When two lines are identical, and only portions of them are visible, we have the perceptually salient property of collinearity. Parallelism is also known to be perceptually salient (eg.,[Abravanel1977]). Given three lines, another viewpoint invariant property arises when the lines all intersect in a single point. A configuration of three intersecting line segments is not typically found in natural stimuli except when the line segments terminate at the same point. In this case, they form a trihedral vertex. Such vertices also play a key role in perception (see, eg.,[Shepard1981

for a discussion). However, trihedral vertices might also be viewed as derived from another minimal viewpoint invariant, identity, because the three endpoints of the line all coincide in the image.

On the other hand, we may define an infinite number of viewpoint invariant properties based on quadruples of lines, because affine invariant representations of quadruples of lines exist (eg.,[Tsai1993]). Once again, a small set of lines gives rise to a small, perceptually salient set of viewpoint invariants while a larger set of lines gives rise to an infinite class of non-perceptually salient viewpoint invariants.

In addition to these properties, it is well known that lines with horizontal or vertical orientation seem to enjoy a special perceptual status (see[Rock1983] for discussion). These properties can also be viewed as minimal viewpoint invariants, when one considers a somewhat different model of image formation. When one allows for general scaled orthographic projection of the world, there can be nothing special about vertical or horizontal lines, because projection allows an arbitrary rotation of the image; no specific orientation of the image is privileged. But let's now consider the subset of scaled orthographic projections that occur when the world is viewed with an upright head. In this case, while the eye can translate arbitrarily in the world, it can rotate only about an axis that is perpendicular to the ground plane. When we limit ourselves to this set of viewpoints, the property of a line being horizontal or vertical becomes a viewpoint invariant one, because horizontal or vertical lines in the world will always appear horizontal and vertical, respectively. But no other angle of a line is invariant; any oblique line in the world may produce a range of angles in the image. Specifically, a line that appears with an angle θ when viewed lying in a fronto-parallel plane will

appear as vertical when viewed after a ninety degree rotation, and may adopt any intermediate angle for intermediate rotations. Therefore the horizontal and vertical may also be viewed as minimal viewpoint invariants. In this case, the property of being a right angle is also a viewpoint invariant one. This is because a scene angle formed by a horizontal and a vertical line will always produce a right angle in an image. No other combination of two lines produce an angle that is invariant to viewpoint, however.

It is also well known that right angles play a special role perceptually. Here, we must distinguish between two separate issues. First, humans have a bias to perceive lines that are not orthogonal in the image as orthogonal in 3-D. This sort of bias really lies outside the scope of our current discussion (but see,[Shepard1981] for extensive discussion that is largely compatible with our discussion here). Here, we focus on the nature of image properties that are perceptually salient, not on the more general issue of how the human visual system selects a particular 3-D percept from among alternatives that are consistent with the 2-D stimulus.

Second is the perceptual salience of angles that are right angles in the image. As[Rock1983] discusses, this bias may be especially clear when the lines that form the right angle are horizontal and vertical, and may be derivative of these properties. It is also interesting to consider a viewing transformation in which a photograph is taken by an upright camera, or a drawing is made from such a viewpoint, but then the photo or drawing is rotated arbitrarily. Horizontal or vertical lines are no longer viewpoint invariant, since when the photograph is rotated a line may take on any orientation. In this case, however, the property of being a right angle is still a viewpoint invariant one, because a horizontal and vertical line in the scene always produce a right angle

in a rotated version of an image taken with an upright camera.

D. The Boundaries of Surfaces

Until now, we have considered properties of isolated geometric features, as if points or lines could just be suspended in a 3-D scene in arbitrary positions. But of course these features arise only in the context of complete scenes. We will now consider how the picture we have drawn changes when the features are produced by the surfaces of objects (we will not consider other sorts of edge features, such as those caused by discontinuities in lighting conditions as when objects cast shadows).

We can divide edges produced by surfaces into two classes: edges produced by discontinuities in objects, and edges produced by discontinuities between objects, because they are on the boundaries of smooth surfaces (see Figure 8). These differ in whether the curve on the surface that produces the edge is viewpoint dependent. A surface discontinuity can produce an edge whenever it is visible. But which part of a smooth object produces a silhouette depends on viewpoint.

Understanding edges caused by surface discontinuities requires little modification of our past discussion. Certain sorts of surface discontinuities reliably cause edges in images when they are visible, such as discontinuities in surface orientation and reflectance. The projections of these curves generally produce edges in images when they are visible (we will not consider the effect that changes in object position relative to a light source may have on the presence or absence of edges). Whenever visible, these curves in 3-D will produce viewpoint invariant properties just as we have described. For example, if a surface contains discontinuities along two parallel line segments, as

for example in a cube, then when they are visible these lines will be parallel in the image.

Edges produced by smooth objects are different. These produce edges when there is a depth discontinuity between the object and its background, or between one part of the object and another. We will call these edges *silhouettes*. The silhouette is produced by a curve on the object that we call the *contour generator*. We will make use of a few basic facts from differential geometry (see eg.,[Koenderink1990]). Each point of a smooth object can be on the contour generator only when the object is seen from at most a 1-D set of viewing directions. This implies that any given point will appear on the contour generator with probability zero. Moreover, two different views of an object will give rise to contour generators that are almost completely different. For a convex object, two distinct contour generators will have two points in common; with objects that contain concavities there will generically be a small finite sets of points shared by two contour generators (we ignore some special cases, such as objects with planar patches that are viewed so that a whole planar patches projects to just a straight line in the image).

The upshot of this is that a view that reveals one contour generator of a general, smooth object actually tells us very little that is certain about the other images this object can produce. For example, given any 2-D silhouette, it is still possible for the object that produced that silhouette to produce any other possible 2-D silhouette (This is discussed further in[Jacobs etal2000]). In general, while we can construct some objects that always produce a particular property, and some objects that never do, we can also construct objects that produce that property from a set of viewpoints

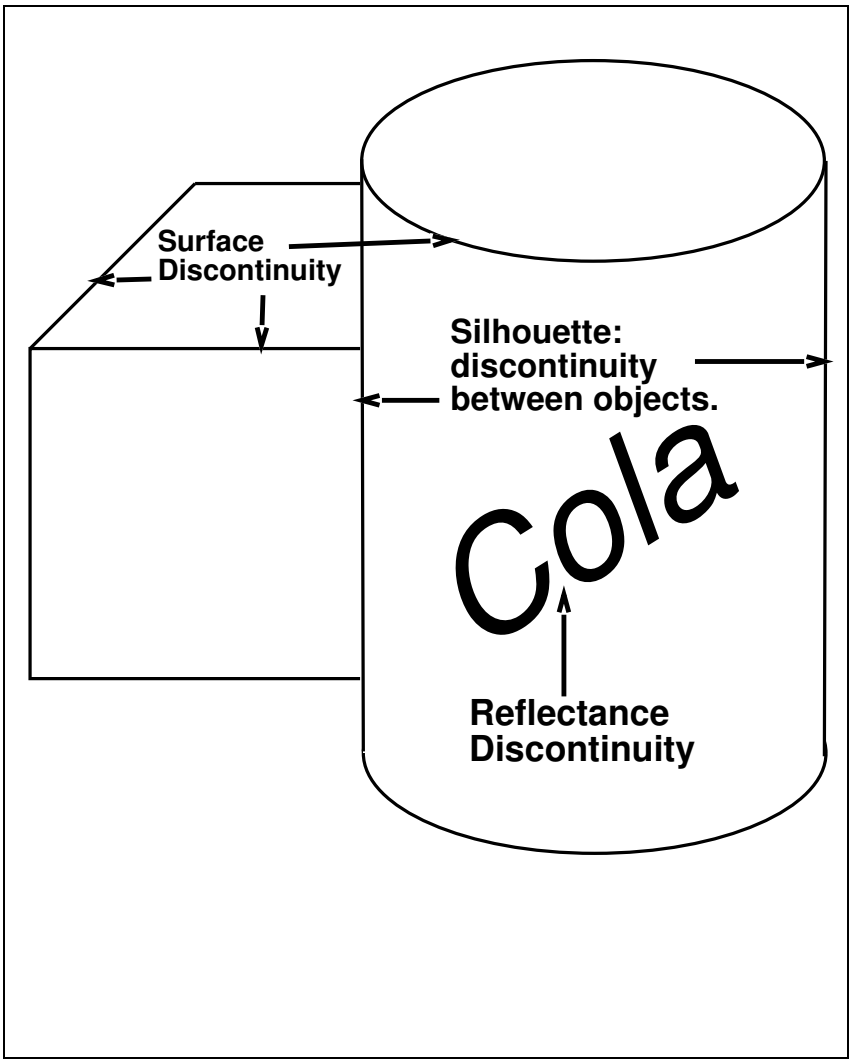

Fig. 8. We consider two fundamentally different types of edges. The position of surface discontinuities, such as discontinuities in orientation or reflectance, are not viewpoint dependent. The same point on the object may produce an edge even as the viewpoint varies. Discontinuities between objects produce the silhouttes of smooth objects. The position on the object that produces these edges is entirely viewpoint dependent.

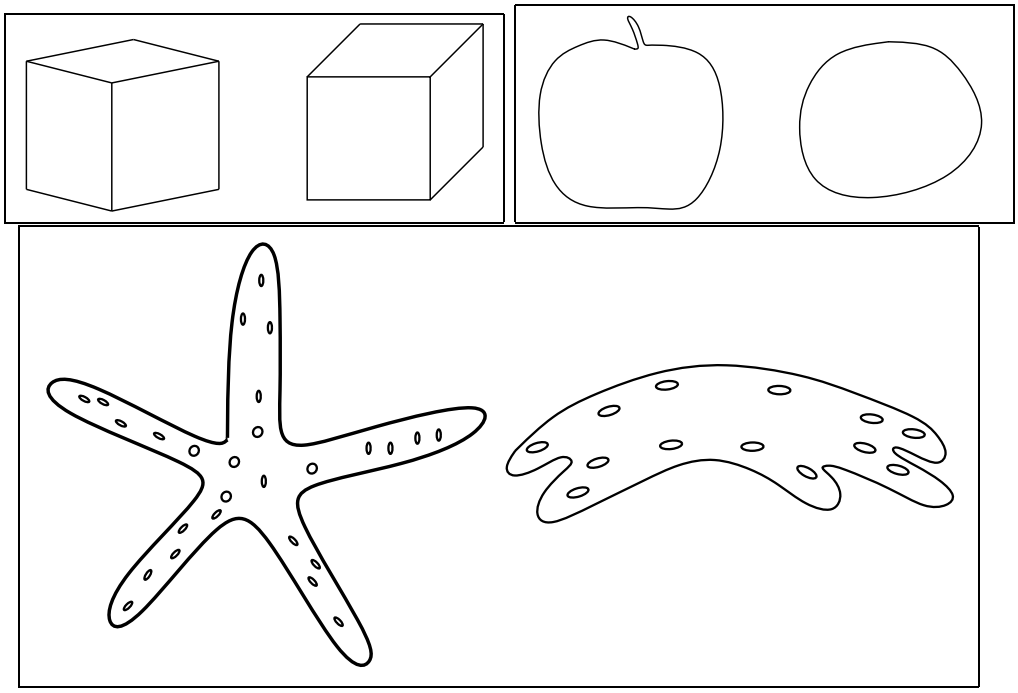

Fig. 9. Convex objects, like the box in the upper left, will produce a convex silhouette when viewed from any direction. Some other objects, like the starfish on the bottom, may produce concavities in their silhouette when viewed from any angle (the legs of the starfish curve downward a bit, so even when viewed end-on, as on the right, they will be somewhat visible). Other objects, like the apple in the upper right, may have convex silhouettes when viewed from some directions, and produce silhouettes with concavities from other directions.

that occur with probability greater than zero but less than one[6].

We will illustrate this for the property of convexity, as applied to the silhouettes of smooth objects (see Figure 9). A convex object will always produce a silhouette that is

[6]Viewpoint invariants may still exist for silhouettes in the following sense. There are some properties that are shared by all silhouettes, that may not be present in edges produced by some other processes. For example, the external silhouette of an object is always closed, while an edge due to a highlight might not be closed. Also,[Koenderink van Doorn1982] point out that the internal silhouettes of smooth objects must end in concavities.

completely convex. Many other objects will never, or almost never, produce a convex silhouette. For example a starfish's silhouette will be convex only when it is viewed end-on, or perhaps never, depending on its exact shape (if the starfish is completely flat, it may look rectangular when viewed end-on; if the legs of the starfish curl a bit, concavities may be visible even from this viewpoint). But plenty of other objects produce silhouettes that are sometimes convex, and sometimes not. The concavities of a pear, for example, are usually visible. But when viewed from below, over a range of directions, the pear will have a nearly circular, convex silhouette. An apple may appear convex when its stem is hidden from view, but non-convex otherwise. Similarly, if an object has a smooth silhouette when viewed from one direction, we cannot conclude that the silhouette will appear smooth from a different direction. Overall, a single silhouette tells us a lot about the shape of its contour generator. For example, a convex silhouette must come from a convex contour generator ([Koenderink1984]). But a single silhouette tells us little if anything that is certain about the shape of the object's other contour generators.

On the other hand, viewpoint invariant properties may arise if we consider restricted classes of objects rather than allowing for completely general objects. For example, if all objects are assumed to be solids of revolution, one can derive invariant properties of their silhouettes. More generally, a series of work has derived invariant properties for certain classes of *generalized cylinders*. A generalized cylinder is the volume swept out by a planar shape as it moves along an axis. Restricted classes of possible generalized cylinders are obtained by limiting the possible shape of the axis, and the way in which the cross-section may change as it sweeps along the axis. For

example, a solid of revolution is a generalized cylinder in which the axis is straight, and the cross-sections are disks that are perpindicular to the axis. But the size of the cross-sections may vary arbitrarily as they sweep along the axis.[Ponce etal1989] discuss invariant properties of a superset of solids of revolutions called straight homogenous generalized cylinders, in which the cross-section need not be a disk, but may have any shape. The shape of the cross-section must remain the same, but its size may vary.,[Zerroug Nevatia1996] discuss generalized cylinders that have curved, but planar axes, and that have cross-sections that are either of constant size, or that are disks of varying size. These authors derive viewpoint properties for these special classes of generalized cylinders, although these invariants are somewhat technical and complex to review here.

Work on generalized cylinders is of considerable interest, because many objects (eg., a vase, a tree trunk, a table leg) belong to these classes. Moreover, many objects used in studies of human vision (eg.,[Biederman Bar1999]) have parts that either belong to these special classes of generalized cylinders, or are planar or are polyhedral with significant surface discontinuities. When we consider specialized classes of objects, we can find a story that is similar to that produced by simple sets of geometric features. For example, when we look at differential properties of silhouettes of rotationally symmetric objects we can show they produce an infinite number of possible viewpoint invariants, but that the classification of points on silhouettes into inflexions, convexity or concavity is still a minimal viewpoint invariant.[7]

[7]A rotationally symmetric object has a line as a central axis, and cross-sections that are disks perpendicular to this axis. Let D stand for one of these disks. Except when viewed from the one

So, for special classes of smooth objects we can again produce many viewpoint invariant properties, but small sets of minimal viewpoint invariants still exist. At the same time, these special classes are quite restricted, compared to all possible objects. Many real objects also do not fall in these classes. Shoes, or feet, for example, do not have cross-sections of unvarying shape. The branches of plants often have a central axis that twists in 3-D space, rather than lying in a plane. Viewpoint invariant

special view in which the axis appears as a single point, D will have two points that appear on the silhouette if they are not occluded. Because of the symmetry of the object, all points on D are equivalent, in that the shape of the object around them must be the same. So even though any specific point on D will usually not lie on the silhouette, we can talk about the properties of whichever point on D happens to appear on the silhouette. So, we can ask whether there is some invariant property of whatever point on the disk is actually visible. It is easy to show, for example that at all points on the same disk the surface of the object has the same 3-D curvature. This means that at all the points on a particular disk, the 3-D surface is either always convex or always hyperbolic (rotationally symmetric objects have no concavities). If, for example, all the points on D are convex in 3-D, then whichever point on D appears on the silhouette will be convex. Similarly if all points on D are hyperbolic (have negative Gaussian curvature) its image will always be concave ([Koenderink1984]). We may also show that if the points on D have zero Gaussian curvature they will always appear on the silhouette as inflexion points, or on lines. Therefore, our decomposition of image points into convex, concave, or inflexions is viewpoint invariant. At the same time, it is easy to show that there are no other viewpoint invariants of the curvature of silhouette points, because scaling the image also scales curvature, and the same reasoning we used in discussing the differential properties of curves still applies. On the other hand, much work has shown that for special classes of generalized cylinders, including rotationally symmetric ones, complete reconstruction from a single silhouette is possible. This implies that any number of viewpoint invariant descriptions may be constructed, using the strategy considered above.

properties as we have defined them do not exist for the silhouettes of completely arbitrary 3-D objects. One possible research strategy for dealing with this is to show that human ability to recognize objects is limited to special classes of objects for which viewpoint invariant properties exist. Such a view seems consistent with a body of work showing strong differences in recognition performance depending on the presence or absence of viewpoint invariant properties. Another possibility is to broaden the notion of viewpoint invariants to understand how the properties of silhouettes can be used for rapid object recognition ([Jacobs etal2000] describes the beginnings of one approach to this).

E. Other Perceptually Salient Properties

We do not attempt to treat all perceptually salient properties in this paper. For example, we have not considered photometric properties, such as the presence of luminance edges or of regions of smoothly varying intensities. Such properties have to do more with invariance to lighting than to viewpoint. However, we feel that a similar analysis can be applied to such properties. For example, it is known that there are no properties of intensity images that are completely invariant to lighting changes ([Moses Ullman1992, Jacobs etal1998]), but also that the presence of an intensity discontinuity is lighting invariant, in a sense similar to that we have used for viewpoint invariance. Moreover, one can show that the presence of an edge is a minimal lighting invariant.

F. Summary

In summary, we have considered a variety of geometric features, including points, lines, curves and surfaces. We have shown that sets of these features give rise to infinite numbers of viewpoint invariant properties that are not in general perceptually salient. However, minimal viewpoint invariant properties of these features are perceptually salient, capturing many of the classic Gestalt properties of shape.

4. Discussion: Why Minimal Features?

There is a long history in psychology of attempts to understand exactly why the Gestalt properties are perceptually salient. So far we have focused on the empirical observation that the set of Gestalt properties is almost identical to the set of minimal viewpoint invariants. This could just be a coincidence, a curiosity with no deeper significance. However, in this section we will argue that there is a natural connection between minimal viewpoint invariants and perceptual salience. Specifically, we argue that it is natural for perceptually salient properties to be ones that are both informative about the world, and easy to use computationally. We draw on work in computational vision to demonstrate that minimal properties are the ones that can best be picked up by the visual system using a feasible amount of computation.

We will lead up to this argument by reviewing past explanations of the salience of Gestalt properties, with two purposes in mind. In Section A we first describe past work that points out the inferential leverage provided by Gestalt properties. Our view of the importance of minimal viewpoint invariants rests on this work. Second, we show that this past work does not predict that salient properties will be the minimal ones,

to make clear why the prevalence of minimality among salient properties requires further explanation. Then in Section B we draw on prior work to argue that minimal properties are linked to computational tractability.

Many authors have provided quite different explanations for the salience of Gestalt properties. In particular, many of these approaches see perceptual organization as the process of finding interesting and relevant relations within a single image, rather than finding features in the image that provide cues about the 3-D structure that produced the image. For example, considerable work has focused on finding 2-D relations that facilitate efficient coding of the image. This is a natural goal for the visual system that may be prior to and quite orthogonal to the problem of using images to infer properties of the world. Such an explanation would seem to have to treat the links we have shown between minimal viewpoint invariance and Gestalt properties as just a coincidence. Other than noting that this seems like a suspicious coincidence, we have nothing to add here to the analysis of these explanations for Gestalt salience. For a recent discussion of these approaches, the reader can refer to.[Van der Helm Leeuwenberg1996] A thorough review of these issues is given in.[Pomerantz Kubovy1986Chater1996] has attempted to reconcile views of perceptual organization as efficient coding with those that characterize it as probabilistic inference by relating them through Kolmogorov complexity.[8]

A second approach along these lines that is more closely related to our current

[8]Chater1996 describes mathematical work that shows that the length of the shortest possible coding of a signal, such as an image, is inversely proportional, up to a constant factor, to the probability of that image arising. It should be noted, however, that the constant factors involved can be quite large (easily on the order of 2^{100}), raising some question about the practical implication of this work.

discussion is work in,[Leyton1986] and,[Palmer1983] which views perceptual organization as a process of describing relations within an image according to a group structure. Palmer, in particular, focuses on organization as the detection of structures within an image that are invariant under subgroups of 2-D similarity transformations. This approach, in which invariance is not used as a way to relate the 2-D image to a 3-D structure, is quite different from the ones we are about to discuss.

In general, approaches that focus on the 2-D relations within images produce very natural explanations for the importance of symmetry and repetition of patterns, since they almost begin with the premise that these structures will be important. It is less clear how well these views can explain the salience of image properties such as convexity and closure, which seem naturally to relate 2-D image properties to prevalent 3-D structures. Of course, views such as those that emphasize efficient coding are not necessarily at odds with views that emphasize inference about 3-D, since both may be goals of the visual system that affect performance.

A. Inferential Leverage

Witkin Tenenbaum1983 provide one discussion of the role of inferential leverage in determining what makes a good feature. They claim that the most perceptually useful image properties will be ones that occur frequently due to scene structure, but only rarely due to a noisy background process. Related views have been expressed by others. For example,[Rock1983] considered many perceptually salient features (such as the collinearity of lines) to be coincidences that the visual system would like to explain with a common cause.

Lowe1985 (see also Binford1981) extended Wiktkin and Tenenbaum's view by pointing out that many perceptually salient non-accidental properties are viewpoint invariant properties. Lowe showed how the invariance of properties such as parallelism and symmetry leads to their also being non-accidental, because 3-D scene structures with these properties always produce images with these properties. In Lowe's view, viewpoint invariant properties will be the ones that provide the most leverage in inferring 3-D scene properties from a 2-D image.

Jepson Richards1992 place Lowe's reasoning on a more general footing. They particularly stress that a feature will be useful only when its prior probability is non-zero. They show that if a feature such as collinearity does not have a substantial prior probability of occurring in scenes, then its presence in an image may not lead to a significant posterior probability that the scene contains collinear points. The accidentalness of non-collinear scene points producing collinear image points may be as likely as the accidentalness of a scene happening to contain collinear points in the first place.

From this perspective, it is natural to argue that the perceptually salient properties are those that are both viewpoint invariant, and occur naturally in the world. Therefore, one might argue that collinearity and the other minimal features we discuss will be useful, since they are likely to arise in real scenes, while more contrived features, such as the property defined by $\alpha_4 + 2\beta_4 = 2$ occur only rarely in the world. Again, from this view it would just be a coincidence that the viewpoint invariant properties that are prevalent in real objects are also minimal. We will argue instead that minimality provides the natural vocabulary with which to search for useful reg-

ularities in the world. We say more about this in the next section.

Overall, we feel that this stress on the inferential leverage of non-accidental properties, and especially of viewpoint invariants, has provided a key component in understanding the importance of these properties in perceptual salience. These authors make convincing arguments for the value of the visual system attending at an early stage to those properties that provide the most information about the scene, and demonstrate the important role that viewpoint invariance can have in providing this information.

The question that we raise is why it should be the case that so many perceptually salient properties are also minimal. While this issue has not been directly confronted previously, Witkin and Tanenbaum and Jepson and Richards make observations that are related to ours concerning minimality.

Witkin and Tanenbaum specifically argued that image features that reflect a lack of change or indicate a repetition of a process will occur reasonably often due to scene structure but rarely due to chance. For example, one can view symmetry as the repetition of a pattern in a scene. Witkin and Tenenbaum use[Ullman1979]'s structure from motion work as another example. Ullman argues that a set of dots moving in an image will be interpreted as a rigid 3-D structure when this is possible, because such an interpretation implies stability in the 3-D structure. Witkin and Tenenbaum attempt to also explain the perceptual salience of such properties as parallelism, collinearity, and good continuation, as properties in which change is minimal between features. That is, parallel lines do not change their angle, collinear points continue in the same direction, smooth curves change direction only slowly.

It is no coincidence that Witkin and Tanenbaum explain many of the same properties by a lack of change that we explain as minimal. Mathematically, minimal properties arise when a set of features contain a degeneracy, meaning that they are not independent. For example, two points that are identical are not independent mathematically, since one point may be described by the other. Three collinear points are not independent because the third may be described as a linear combination of the first two; this is not true when the third point is not collinear to the others. While we have avoided detailed mathematical derivation, it can be shown that the viewpoint invariance of identical or collinear points is a direct consequence of their non-independence. This non-independence can also be interpreted, as Witkin and Tanenbaum do, as a lack of change. The fact that a third collinear point is a linear combination of the first two can equally be regarded as a constancy of direction. Hence for some perceptually salient features, minimality and the absence of change can be regarded as equivalent properties.

However, minimality and lack of change are not equivalent in all cases, as in for example convexity, or our treatment of horizontal and vertical directions as minimal viewpoint invariants, and in the case of symmetry, which we will discuss below. Moreover, as we will discuss in the next section, minimality leads to a different sort of explanation for perceptual salience, rooted in computational considerations rather than just concern for which properties provide the most leverage in inferring scene properties.

Jepson and Richards make use of the notion of *codimension*, in explaining the perceptual salience of image features. The codimension of an image property is essen-

tially the dimension of all possible images, minus the dimension of the set of images with that property. For example, the property that three image points are collinear has codimension one. To see this, note that the dimension of the possible configurations of three image points is six, a dimension for each coordinate of each point. The dimension of the set of configurations of the points that are collinear is five, since the third point has only one, not two remaining degrees of freedom. Jepson and Richards suggest that useful features will have non-zero codimension in the scene, and in corresponding images. This is true of all perceptually salient viewpoint invariants we have considered. It is also true of all non-salient viewpoint invariants we consider. For example, planar point sets with the property defined by $\alpha_4 + 2\beta_4 = 2$ have codimension one, as do the images they produce. Therefore, Jepson and Richards' observations on codimension, while of interest, are not equivalent to our discussion of the minimal nature of perceptually salient viewpoint invariants, nor does codimension alone provide an alternative explanation for the perceptual salience of minimal viewpoint invariants.

B. Minimal Properties

We now consider reasons why the minimality of features may play a special role in early vision. We will point out that the complexity of detecting and using properties seems to depend on the number of features on which the property is based. This is relevant because it indicates that visual cortex may be able to build representations out of minimal features using fewer neurons than more complex features might require.

The notion of computational complexity is well-developed for parallel systems

(eg., Leigton1992) and one might hope in principle to use it to prove that identifying and using non-minimal non-accidental properties (NAPs) will take inherently more neurons or time than it will to handle minimal NAPs. In practice, however, such proofs are well beyond the reach of existing techniques. Therefore, our approach will be more heuristic. We will look at the methods that have been developed for identifying NAPs, focusing on the ones developed to model neural architectures. Then we will see how the complexity of these approaches would grow if we applied them to non-minimal properties.

A number of computational approaches to computing good continuation have been proposed. They are applied to modelling human ability to detect curves in noisy environments, including phenomena of illusory contour formation. The idea is that the image provides some fragmentary information about the position and orientation of contours, which are completed in a way that favors smooth curves. In one class of algorithms, each position and orientation is explicitly represented by a separate unit. Then this unit gathers evidence that a contour passes through the location it represents from a large surrounding neighborhood. This is done with a convolution, that is, by a linear combination of the activity in the surrounding neighborhood. Algorithms that fall into this general framework include Grossberg Mingolla1985, Heitger Von der Heydt1993, Guy Medioni1996 and. Williams Jacobs1 Typically, this linear convolution is then followed by a non-linear step that may, for example, threshold, perform non-maximum suppression or combine information coming from different directions.

These methods produce smoothly interpolated curves, where smoothness is a property of curvature. In order to gather evidence for smoothness with a single linear

operation, it is important to explicitly represent position and orientation. This allows the convolution to weight the extent to which a contour at one position and orientation supports that at another position and orientation. The convolution weights can be based on a difference in orientation, and so can capture properties of curvature. To make use of differential properties of curves that are higher order than curvature with such a mechanism would require computational units that explicitly represent properties that are higher order than orientation. For example, existing systems can find contour completions that minimize total curvature (eg.,[Williams Jacobs1997a]). To minimize changes in curvature with the same mechanisms would require a system that explicitly represents each position, orientation, and curvature. This would require at least an order of magnitude more neurons. There is a clear computational advantage in avoiding such representations if possible.

A number of other models have been proposed that rely on iterative local interactions between computational units, rather than large scale convolutions ([Parent Zucker1989, Sha'ashua Ullman1
These systems enforce smoothness of curvature by the interaction between adjacent units that explicitly represent position and orientation, upon which curvature is based ([Parent Zucker1989] also explicitly represent curvature). Related, very interesting work can be found in.[Krieger Zetzsche1996] This shows that while first order properties of contours can be computed using linear filters, detection of second order properties, based on curvature, require more complex, non-linear filters that combine the results of linear systems in a non-linear way. In general, sensitivity to minimal invariants seems to require simpler neural hardware than would be needed to handle properties based on more features.

Some of these models can be mapped into general framework of Markov models, which have well studied computational properties. A Markov model is a probabilistic model in which a value is assumed to be independent of all other values except for those in a small neighborhood. For example, a Markov model of image intensities might encode the expectation that regions will tend to have uniform intensities, with occasional discontinuities between neighboring regions (eg., Geman and Geman, 1984). This can be encoded by assuming that a pixel is likely to have an intensity that is similar to that of its neighbor, with a small probability of a large change in intensity. A Markov model of contours might similarly capture our expectation that the contours of objects tend to be smooth (Mumford, 1994), perhaps with occasional discontinuities (Thornber and Williams, 1997). These prior expectations are then combined with the actual information present in an image to produce the most likely interpretation of an image.

The complexity of learning or computing with a Markov model seems to depend critically on the size of the *neighborhood* used. As we have seen, differential properties of curves can be expressed in terms of the properties of the points in a local neighborhood about the curve. Markov models of curves that are sensitive to minimal properties, such as corners or smoothness, require smaller neighborhoods than models that are sensitive to higher order properties, such as non-minimal NAPs.

We can build Markov models using larger neighborhoods if we desire. However, in practice computer vision researchers try to use the smallest possible Markov models, since the computational complexity of using these models grows with the size of these neighborhoods. Markov models with larger neighborhoods have more parame-

ters. For a visual system to learn an accurate model with more parameters, it must have more data. Moreover, the amount of data needed can grow exponentially with the number of parameters, ie., with the size of the neighborhoods used. These considerations show up in neurally inspired models of illusory contour formation. Most specifically,[Williams Jacobs1997b] show how to use a Markov model of contours for illusory contour formation. As mentioned above, in this and other models, finding contours that minimize curvature requires explicit representation of position and orientation, and using higher order Markov models would require the additional explicit representation of curvature.

This argument is also in line with the psychophysical results of.[Feldman1997] Feldman studies human judgements about the curvilinearity of dot patterns. His results show that human analysis of contours is based on small sets of adjacent points. People seem to make judgments about contours based on the properties of quadruples of points. Judgements about contours with more points can be interpreted as the combination of judgements about adjacent quadruples of points. This is what one would expect if subjects implicitly analyze contours using a low-order Markov model.

We are therefore suggesting that the visual system is sensitive to properties that provide inferential leverage and that can be detected efficiently with neural hardware. Minimal viewpoint invariants are the most extreme examples of such properties. These properties are local, in the sense that they can be based on local neighborhoods that involve a small number of features, or a small portion of the image. This does not mean that global properties are not important. Rather it means that the visual system will emphasize global properties that can be constructed from local properties.

Markov models provide one well-developed mechanism that can produce this behavior.[Mumford1994] describes how to capture good continuation this way. Another simple example is closure. A Markov model can enforce closure because end points can be detected locally, and this information can be propagated globally.[Zhu1999] discusses how to capture other Gestalt properties using Markov models. Below, we will discuss the work of Wagemans and his collaborators, which shows how symmetry can be detected, not with a Markov model, but with a model that detects minimal NAPs and then combines them. Overall, it is our contention that in this way, minimal NAPs provide the basic primitives out of which early vision can construct interpretations of images. And we suggest that the reason for this is both that they provide inferential leverage and they lead to tractable computations.

C. Beyond minimal viewpoint invariance

Minimal viewpoint invariants seem to play a special role in perception. But there is not a one-to-one correspondence between minimal viewpoint invariants and perceptual salience, nor does our approach predict that there should be. We claim that the visual system will be especially sensitive to properties that provide inferential leverage and that can be computed tractably by the brain. Minimal viewpoint invariants will be the most extreme, but not the only examples of such properties. Also, it turns out that they can help us to understand the value of a number of other, related properties.

First, we note that minimal viewpoint invariants can be viewed as the limiting case of other salient properties. For example, proximity is a classic Gestalt grouping cue, as dots that are nearby are more readily grouped together than distant dots.

Identity can be seen as the limiting case of proximity. Similarly, collinearity can be seen as the limiting case of smoothness. And we have pointed out that closure can be considered a derivative property of identity. A contour is closed when its end points are identical. In practice, psychophysics demonstrating the value of closure has used stimuli in which contours contain gaps ([Elder Zucker1994, Kovacs Julesz1993]). Closure is only the limiting case of these stimuli. Of course, other properties, such as parallelism, are also never perfectly present in an image, due to noise and limitations on sensing. So if our discussion of minimal viewpoint invariants is to be relevant to a wide range of conditions we need to show that it has implications for properties that are possessed to some degree. We will discuss proximity as a canonical example of this.

If two points are close together in a scene, they need not always appear close together in an image; when we get close to the points, they can be widely separated in the image. In fact, our proof that identity is the only viewpoint invariant property of two points shows that proximity cannot be a viewpoint invariant. But, while proximity is not a viewpoint invariant, it's presence in the image does provide a lot of probabilistic information about the 3-D scene. Points that are nearby in the scene are usually much closer to each other in the image than are points that are widely separated in the scene. Therefore, while not providing as much information as identity, proximity is still a powerful source of inferential leverage.

Properties like proximity, which are preserved with some small degradation over a large range of viewpoints are called *quasi-invariants* in the computer vision literature (for some discussion, see Zerroug and Nevatia, 1996, and Binford and Levitt, 1993). We can show formally that any viewpoint invariant property will have "near" it, a

quasi-invariant property. That is, if a property is perfectly preserved under viewpoint changes, than a scene that is close to having this property will produce images that are close to having that property from a wide range of viewpoints. Proximity is the quasi-invariant property of being nearly identical. As such, it is based on a small number of image features, and provides considerable inferential leverage. Therefore, the same arguments that indicate that minimal viewpoint invariants may be useful to the visual system will also apply to "minimal" quasi-invariants, such as proximity. We may also note, that as the distance between points increases, it becomes less of an indication of any specific distance between the scene points. This can explain why a reduction in proximity can lead to a continuous reduction in salience.

Essentially the same reasoning applies to other "near" invariants, such as a tolerance for small amounts of curvature, but not large amounts ([Field etal1993]). When a curvature is small, the curve is nearly straight, that is, it nearly has a viewpoint invariant based on collinearity. This is also a quasi-invariant, since a part of a curve with small curvature, will nearly always appear in an image with little curvature.

A second possible source for the perceptual salience of properties that are near invariants is that they may reflect the presence of an invariant property that is corrupted by noise, or small perspective effects. For example, if two lines are nearly parallel in an image, they may be produced by parallel scene lines, with sensor noise accounting for the deviation from parallel. Or they may converge slightly due to small amounts of perspective distortion. A contour with small gaps in it may come from a closed contour that is partly occluded or fragmented due to noise.

The fact that perceptual salience is not all or nothing can also be related to

neurophysiology. Neurons do not show an all or nothing response to a specific property, such as a specific orientation, but rather show a graded response. We can see that the informativeness of quasi-invariants makes a graded response to an invariant a valuable way of gathering information about the 3-D scene. Presumably, a network that adapts the statistical relations between 2-D image properties and a 3-D scene will display a similar,graded response ([Hummel Biederman1992] gives one example of such a system).

Both the informativeness of quasi-invariants, and the fact that image properties that are near invariants may be true invariants corrupted by noise mean that it makes sense for a perceptual system to lump together invariant properties with properties that are close to them. If this biases the visual system to sometimes, but not always perceive a nearly closed contour as closed, then this explains why a closed contour will not pop out of a field of slightly open contours, because some slightly open contours are also perceived as closed. This explanation has essentially been put forward in.[Williams Julesz1992]

Proximity and smooth continuation are related to minimal properties, but they are not truly viewpoint invariant. Symmetry is a key Gestalt property that is not truly minimal. And it also exhibits gradations in salience depending on the type of symmetry. While we do not attempt a complete explanation of all the data on symmetry, we wish now to indicate how symmetry can fit into, and also stretch, the picture we have drawn above.

There are a number of different forms of symmetry; we will consider *mirror symmetry*, in which points in a 2-D plane are reflected about a line, and *skew symmetry*.

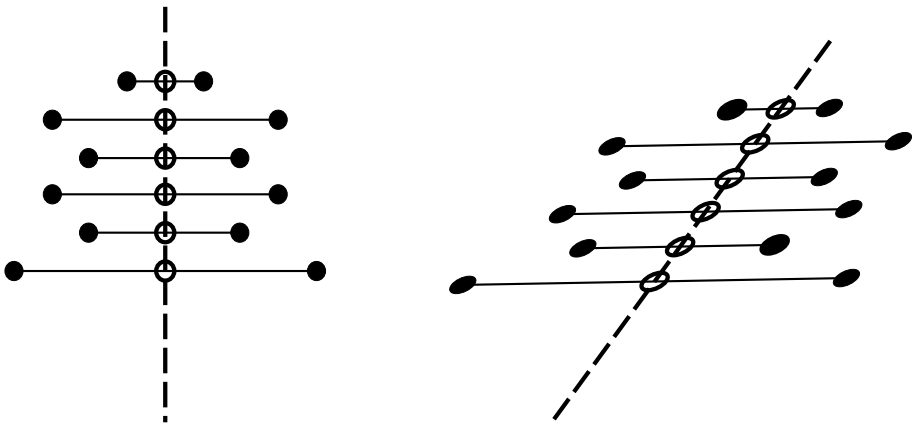

Fig. 10. On the left, a set of symmetric points (closed circles). On the right, the points are *skewed symmetric*, that is, they are a scaled orthographic projection of the points on the left. Points are skewed symmetric points only when the midpoints (shown as open circles) of the corresponding points are collinear (they are shown lying on a dashed line). They are symmetric when, additionally, the line formed by the midpoints is orthogonal to the lines joining corresponding points.

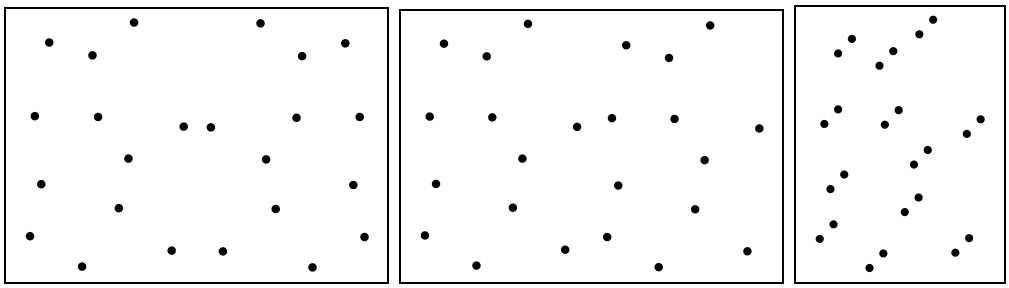

Fig. 11. On the left, mirror symmetry, in the middle, translational symmetry, on the right, a Glass pattern. The symmetry of the middle figure is usually harder to detect.

We say that a set of points is skew symmetric if and only if it could be the scaled orthographic projection of a set of mirror symmetric points. Suppose that points $\mathbf{p}_{1,1}, ...\mathbf{p}_{n,1}, \mathbf{p}_{1,2}, ...\mathbf{p}_{n,2}$ are symmetric, with the set $\mathbf{p}_{1,2}...\mathbf{p}_{n,2}$ being the reflection of the set $\mathbf{p}_{1,1}...\mathbf{p}_{n,1}$. Then we can show that the projection of these points, $\mathbf{q}_{i,j}$ will be skew symmetric if and only if all the points midpoint between $\mathbf{q}_{i,1}$ and $\mathbf{q}_{i,2}$, for all i, are collinear (see Figure 10).

Skew symmetry is a viewpoint invariant property, by its definition. Mirror symmetry is not viewpoint invariant for general viewing conditions, although it is when one restricts the relative rotation between the object and viewer to be about the axis of symmetry. Mirror symmetry is preserved by projection when a person looking parallel to the ground plane views an object standing upright with its symmetry axis perpendicular to the ground plane. An example of this is when people look at other people or animals standing upright.

At the least, though, skew symmetry seems a more general viewpoint invariant than mirror symmetry. But[Wagemans etal1992] demonstrate that skew symmetry is less salient, and this has been used as an argument against invariance alone as a predictor of perceptual salience ([Van der Helm Leeuwenberg1996]). Moreover, neither mirror nor skew symmetry are minimal properties. All triples of points are skew symmetric, and in general both symmetries are really only meaningful properties of sets of four or more points.

[Wagemans1999] has argued that none-the-less, the relative salience of these and other symmetric properties can be understood in terms of the computational ease of using these properties. His argument fits well into our overall view of perceptual salience,

and we repeat it now with slight variations to fit our framework.

According to the model of,[Wagemans etal1991] the visual system detects symmetry in a dot pattern by first grouping nearby dots more or less randomly, forming virtual lines between them. Parallelism is noticed between some of these virtual lines, which is a cue to symmetry. When the midpoints of parallel virtual lines are found to be collinear, this determines that the points that give rise to these lines are symmetric. Moreover, regularities in the symmetry can be used to help propagate initial matches. For example, given one virtual line, another line that joins symmetric points must be parallel.[Wagemans etal1991] also stress that in the case of mirror symmetry there is more information that allows us to tell whether two sets of lines are part of a symmetric pattern, since there is also the constraint that the line connecting their centers be orthogonal to the parallel lines. Wagemans argues that the more local constraints there are that indicate whether a small set of points is part of a symmetric pattern, the easier it will be for the visual system to find the symmetry in a figure.

Note several key things about this explanation. First, symmetry is determined using as building blocks the minimal viewpoint invariants of parallelism and collinearity, and the minimal quasi-invariant of proximity. So although symmetry is not a minimal viewpoint invariant, it can still be efficiently computed by the visual system using minimal viewpoint invariants. This is reminiscent of Markov models that build up non-local properties by the conjunction of overlapping local properties. Also, the fact that some forms of symmetry, such as translational symmetry are less salient than mirror symmetry or Glass patterns is explained because translational symmetry lacks the minimal quasi-invariant proximity, making it harder to compute (see Fig-

ure 11). This suggests that the presence of *minimal* properties is essential in making viewpoint invariant properties usable.

As a final note, we want to stress that while it appears that viewpoint invariance plays a key role in perceptual salience, there are also significant limitations of minimal viewpoint invariant properties as representations of images. It is noteworthy that all the minimal viewpoint invariants we have discussed are produced by features that lie in only a zero-, one- or two-dimensional subspace of a three-dimensional scene. Collinearity is a feature of points lying along a line, while co-termination is in fact a zero-dimensional property based on two points being exactly the same. But also, parallelism is a property only of coplanar lines; convexity as we describe it is a property of points or portions of curves that lie in a plane; the smoothness of curves or their corners are also determined by portions of curves that lie locally in a plane. All minimal viewpoint invariants must be computed from features that at most lie in a 2-D plane in the scene. Sets of scene features that are fully three-dimensional do not give rise to such properties.

The significance of this fact for human perception is that minimal viewpoint invariants can only capture planar aspects of a 3-D shape, they cannot describe the relationships between scene features that do not lie in the same plane. They can therefore tell us only some things about the 3-D structure of a scene. For example, when we use parallelism between image lines to infer that the lines are parallel in 3-D, we have also inferred that the lines are coplanar in the scene. We have inferred something about the structure of a 3-D scene, but only something about a planar component of that 3-D scene. There are other important scene properties that minimal viewpoint

invariants cannot help us to infer, such as the relative size of the different parts of an object, or the orthogonality of the edges in a trihedral vertex. To understand these 3-D properties in a scene, it seems that we must make use of image properties that are more metric, and not viewpoint invariant, such as the relative size of different things in the image, or the angles formed at corners in an image.

Biederman1987 has argued that viewpoint invariant properties (the ones we have shown to be minimal) are used to solve the indexing problem in visual object recognition. This means that qualitative properties of objects encoded in viewpoint invariants are used to trigger the appropriate class of object to match an image. The classes of objects that can be identified with this model seem to correspond to what Rosch etal1976 call *basic* level categories. At the same time, it is clear that metric properties must play a significant role in object recognition. At the least, metric properties may be the only ones that distinguish two objects that belong to the same basic category. Metric properties also play a role in models of object recognition, such as, Hummel Biederman1992, Poggio Edelman1990, Lowe1985 and. Ullman1989

Unfortunately, understanding the correct use of metric properties in models of human object recognition is probably one of the most daunting problems in the study of human vision. Metric properties have been used in template matching schemes (Roberts1965 is probably the earliest example for 3-D vision) but these methods seem directly applicable only to the problem of recognizing a rigid object that is precisely known. Human recognition seems to be more a judgement of similarity, in which distortions due to articulations of parts, non-rigid deformations, and subtle variations are understood and tolerated. For example, when we recognize a person, we must be

able to cope with changes in the position of their limbs and facial expression, changes in their hair style, and changes due to aging. A true theory of object recognition seems to call for a model that can judge how similar a new image is to one's past experience of an object, and whether the apparent differences are of a kind that can be disregarded, or that indicate the presence of a new object.

No existing model of recognition comes close to dealing with all these issues. But some models address parts of this problem in interesting ways that may be relevant to a more complete understanding of how we use metric properties. For example,[Ullman1989] suggests that local affine distortions in an object are disregarded.[Todd etal1996] suggest that global affine distortions are more generally disregarded, or of less consequence, in perception.[Hummel Biederman1992] show how metric properties can be integrated into a single system, along with viewpoint invariant properties.[Edelman1998] provides an extensive discussion of similarity judgements, and how they can be built up from experience with an object.[Goldmeier1972] describes some early work on visual similarity that is still quite relevant.

It is our view that minimal viewpoint invariants are useful because they capture generic properties that are relevant to most objects. These properties can be used to trigger more specific models of objects. It is not clear when and if viewpoint invariants will continue to be useful in later visual processes, after more specific models have been activated, but this is a subject of our ongoing work.

5. Conclusion

The main goal of this paper has been to make the empirical observation that most perceptually salient properties are minimal viewpoint invariants. These properties include closure, connectedness, collinearity, parallelism, corners, trihedral vertices, convexity, contour smoothness, and can include the salience of vertical and horizontal lines. Moreover, all minimal viewpoint invariants for the sets of features we have studied are perceptually salient. We have also sketched how symmetry, and "near invariant" properties such as proximity can be included in this framework. The role of viewpoint invariance in vision has previously been much discussed, so our focus has been on showing that *minimal* viewpoint invariants play a role in perception. We feel that our observations have two potential implications for the study of human vision: first to clarify and strengthen our understanding of how viewpoint invariance may be important; second to lend support for explanations of perception that stress that the representations produced by the visual system are not solely based on the potential information these representations make explicit, but also by the ease with which these representations can be computed.

While the significance of viewpoint invariance in vision has been extensively discussed, it has not previously been clear to what extent viewpoint invariance alone can predict which features the visual system finds most important. On one hand, we have shown that viewpoint invariance by itself is not an adequate explanation for perceptual salience. But we have also shown that when combined with minimality, viewpoint invariance becomes a very good predictor of salience. We feel that therefore

our observations patch a significant hole in previous work on the role of viewpoint invariance.

The fact that key image properties are composed of the smallest possible number of features is striking, because most computational approaches to perceptual organization have been driven to use such image descriptions for computational reasons., Wagemans1999 has argued that such computational considerations can help explain the varying degree of salience of different forms of symmetry. We point out that computational considerations seem relevant to a large number of Gestalt properties. Representations that express global properties of an image solely in terms of combinations of local properties cannot capture everything, but they offer a good compromise of expressiveness and computational ease.

Theoretical approaches to explaining human perceptual organization have generally been either motivated by information theory, Bayesian or probabilistic reasoning, or mechanistic explanations. Information theoretic approaches stress the importance of efficient coding of images, and propose that the visual system is sensitive to properties that lead to efficient coding (eg., Attneave1954, Garner1962, Leeuwenberg1971, Van der Helm Leeuwenberg1996). Probabilistic approaches point out that the goal of vision is to figure out what is out there in the world. They claim that the visual system will be sensitive to properties that provide the most information about the world, resting on a foundation that is either implicitly or explicitly probabilistic (eg., Witkin Tenenbaum1983, Rock1983, Lowe1985, Jepson Richards1992). Mechanistic approaches examine the properties of neurons in the visual system and attempt to explain perceptual organization as a by-product of these properties; these approaches are less teleological, attempting more to figure out what is happening than in

predicting what is being computed from first principles (eg., Heitger Von der Heydt1993, Grossberg Mingolla1985). To the computer scientist, these approaches all seem important, but they seem to leave out the equally important consideration of computational efficiency. In particular, a representation might be ideal for purposes of coding or information theory, and still the brain might fail to produce it because it is too difficult to compute. That is, computing the ideal image properties might require too much time, too many neurons, or too complex a set of interconnections between neurons. Or, it might be practical to compute an ideal representation, but not worth the effort. Another representation that is less than ideal, but good enough, might require fewer resources to compute. We claim that the importance of minimality in predicting what properties will be salient is evidence for the importance of computational considerations in understanding human vision.

Acknowledgment

This paper has benefited substantially from several suggestions by Ronen Basri and Irving Biederman. Zili Liu, Bosco Tjan, and Lance Williams have also provided significant assistance.

References

[Abravanel1977] Abravanel, E. (1977). "The figural simplicity of parallel lines." *Child Development*, *48*, 708-710.

[Attneave1954] Attneave, F. (1954). "Some informational aspects of visual perception." *Psychological Review*, *68*, 183-193.

[Basri Moses1999] Basri, R. Moses, Y. (1999). "When is it possible to identify 3D objects from single images using class constraints?" *International Journal of Computer Vision*, *33*(2), 1-22.

[Biederman1987] Biederman, I. (1987). "Recognition-by-components: a theory of human image understanding." *Psychological Review*, *94*(2), 115-147.

[Biederman Bar1999] Biederman, I. Bar, M. (1999). "One-shot viewpoint invariance in matching novel objects." *Vision Research*, *39*, 2885-2899.

[Biederman Gerhardstein1993] Biederman, I. Gerhardstein, P. (1993). "Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance." *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 1162-1182.

[Biederman Gerhardstein1995] Biederman, I. Gerhardstein, P. (1995). "Viewpoint-dependent mechanisms in visual object recognition: reply to Tarr and Bülthoff." *Journal of Experimental Psychology: Human Perception and Performance*, *21*(6), 1506-1514.

[Binford1981] Binford, T. (1981). "Inferring surfaces from images." *Artificial Intelligence*, *17*, 205-244.

[Burns etal1992] Burns, J., Weiss, R. Riseman, E. (1992). "The non-existence of general-case view-invariants." In J. Mundy A. Zisserman (editors, *Geometric invariance in computer vision.* Cambridge, MA: MIT Press.

[Chater1996] Chater, N. (1996). "Reconciling simplicity and likelihood principles in perceptual organization." *Psychology Review*, *103*(3), 566-581.

[Clemens Jacobs1991] Clemens, D. Jacobs, D. (1991). "Space and time bounds on model indexing." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(10), 1007-1018.

[Corballis1988] Corballis, M. (1988). "Recognition of disoriented shapes." *Psychological Review*, *95*, 115-123.

[Cutting1983] Cutting, J. (1983). "Observations: four assumptions about invariance in perception." *Journal of Experimental Psychology: Human Perception and Performance*, *9*(2), 310-317.

[Cyganski etal1987] Cyganski, D., Orr, J., Cott, T. Dodson, R. (1987). "Development, implementation, testing, and application of an affine transform invariant curvature function." In *Proceedings of the first international conference on computer vision.*

[Edelman1998] Edelman, S. (1998). "Representation is representation of similarities." *Behavioral and Brain Sciences*, *21*, 449-498.

[Elder Zucker1994] Elder, J. Zucker, S. (1994). "A measure of closure." *Vision Research*, *34*(24), 3361-3369.

[Feldman1997] Feldman, J. (1997). "Curvilinearity, covariance, and regularity in perceptual groups." *Vision Research*, *37*(20), 2835-2848.

[Field etal1993] Field, D., Hayes, A. Hess, R. (1993). "Contour integration by the human visual system: evidence for a local "association field"." *Vision Research*, *33*(2), 173-193.

[Forsyth etal1992] Forsyth, D., Mundy, J., Zisserman, A. Rothwell, C. (1992). "Recog-

nising rotationally symmetric surfaces from their outlines." In *European conf. on comp. vis.*

[Garner1962] Garner, W. (1962). *Uncertainty and structure as psychological concepts.* New York, NY: Wiley.

[Goldmeier1972] Goldmeier, E. (1972). *Similarity in visually perceived forms.* New York, NY: International Universities Press.

[Grossberg Mingolla1985] Grossberg, S. Mingolla, E. (1985). "Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading." *Psychological Review*, *92*(2), 173-211.

[Guy Medioni1996] Guy, G. Medioni, G. (1996). "Inferring global perceptual contours from local features." *International Journal of Computer Vision*, *20*(1/2), 113–133.

[Hayward Tarr1997] Hayward, W. Tarr, M. (1997). "Testing conditions for viewpoint invariance in object recognition." *Journal of Experimental Psychology: Human Perception and Performance*, *23*(5), 1511-1521.

[Heitger Von der Heydt1993] Heitger, R. Von der Heydt, R. (1993). "A computational model of neural contour processing, figure-ground segregation and illusory contours." In *International conference on computer vision.*

[Hoffman Richards1984] Hoffman, D. Richards, W. (1984). "Parts of recognition." In S. Pinker (editor, *Visual cognition.* Cambridge, MA: MIT Press.

[Hummel Biederman1992] Hummel, J. Biederman, I. (1992). "Dynamic binding in a neural network for shape recognition." *Psychological Review*, *99*(3), 480-517.

[Jacobs1996b] Jacobs, D. (1996b). "The space requirements of indexing under per-

spective projection." *IEEE Trans. on Pattern Analysis and Machine Intelligence*, *18*(3), 330–333.

[Jacobs1997] Jacobs, D. (1997). "Matching 3-D models to 2-D images." *International Journal of Computer Vision*, *21*(1/2), 123–153.

[Jacobs2000] Jacobs, D. (2000). "What makes viewpoint invariant properties perceptually salient?: A computational perspective." In K. Boyer S. Sarkar (editors, *Perceptual organization for artificial vision systems.* Kluwer Academic Publishers.

[Jacobs etal1998] Jacobs, D., Belhumeur, P. Basri, R. (1998). "Comparing images under variable illumination." In *IEEE conference on computer vision and pattern recognition.*

[Jacobs etal2000] Jacobs, D., Belhumeur, P. Jermyn, I. (2000). *Judging whether multiple silhouettes can come from the same object.* In preparation.

[Jepson Richards1992] Jepson, A. Richards, W. (1992). *What makes a good feature?* (Memo #1356). MIT AI Lab.

[Jolicoeur Kosslyn1983] Jolicoeur, P. Kosslyn, S. (1983). "Coordinate systems in the long-term memory representation of three-dimensional shapes." *Cognitive Psychology*, *15*, 301-345.

[Julesz1993] Julesz, B. (1993). "Subjective contours in early vision and beyond." In I. Cox, P. Hansen B. Julesz (editors, *Partitioning data sets.* American Mathematical Society.

[Kanade1981] Kanade, T. (1981). "Recovery of the three-dimensional shape of an object from a single view." *Artificial Intelligence*, *17*, 409-460.

[Koenderink1984] Koenderink, J. (1984). "What does the occluding contour tell us about solid shape?" *Perception*, *13*, 321–330.

[Koenderink1990] Koenderink, J. (1990). *Solid shape.* Cambridge, MA: MIT Press.

[Koenderink van Doorn1982] Koenderink, J. van Doorn, A. (1982). "The shape of smooth objects and the way contours end." *Perception*, *11*, 129-137.

[Koffka1935/1963] Koffka, K. (1963). *Principles of gestalt psychology.* New York, NY: Harcourt, Brace and World. (1935)

[Kovacs Julesz1993] Kovacs, I. Julesz, B. (1993). "A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation." *Proc. Nat. Acad. Sci. USA*, *90*, 7495-7497.

[Krieger Zetzsche1996] Krieger, G. Zetzsche, C. (1996). "Nonlinear image operators for the evaluation of local instrinsic dimensionality." *IEEE Transactions on Image Processing*, *5*(6), 1026-1042.

[Kurbat1994] Kurbat, M. (1994). "Structural description theories: Is RBC/JIM a general-purpose theory of human entry-level object recognition?" *Perception*, *23*, 1339–1368.

[Leeuwenberg1971] Leeuwenberg, E. (1971). "A perceptual coding language for visual and auditory patterns." *American Journal of Psychology*, *84*, 307-349.

[Leigton1992] Leigton, F. (1992). *Introduction to parallel algorithms and architectures: arrays, trees, hypercubes.* M. Kaufmann.

[Leyton1986] Leyton, M. (1986). "A theory of information structure II. a theory of perceptual organization." *Journal of Mathematical Psychology*, *30*, 257-305.

[Lowe1985] Lowe, D. (1985). *Perceptual organization and visual recognition.* The Netherlands: Kluwer Academic Publishers.

[Moses Ullman1992] Moses, Y. Ullman, S. (1992). "Limitations of non model-based recognition schemes." In *Second european conference on computer vision.*

[Mumford1994] Mumford, D. (1994). "Elastica and computer vision." In C. Bajaj (editor, *Algebraic geometry and its applications.* New York, NY: Springer-Verlag.

[Palmer1983] Palmer, S. (1983). "The psychology of perceptual organization: a transformational approach." In J. Beck, B. Hope A. Rosenfeld (editors, *Human and machine vision.* New York, NY: Academic Press.

[Parent Zucker1989] Parent, P. Zucker, S. (1989). "Trace inference, curvature consistency and curve detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(8), 823-839.

[Poggio Edelman1990] Poggio, T. Edelman, S. (1990). "A network that learns to recognize 3D objects." *Nature*, *343*, 263-266.

[Pomerantz Kubovy1986] Pomerantz, J. Kubovy, M. (1986). "Theoretical approaches to perceptual organization." In K. Boff, L. Kaufmann J. Thomas (editors, *Handbook of perception and human performance: Vol. II. cognitive processes and performance* (36.1-36.46). New York, NY: Wiley.

[Ponce etal1989] Ponce, J., Chelberg, D. Mann, W. (1989). "Invariant properties of straight homogeneous generalized cylinders and their contours." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(9), 951-966.

[Roberts1965] Roberts, L. (1965). "Machine perception of three-dimensional solids."

In J. Tippett et al. (editors, *Optical and electro-optical information processing.* Cambridge, MA: MIT Press.

[Rock1983] Rock, I. (1983). *The logic of perception.* Cambridge, MA: MIT Press.

[Rock DiVita1987] Rock, I. DiVita, J. (1987). "A case of viewer-centered object perception." *Cognitive Psychology*, *19*, 280-293.

[Rosch etal1976] Rosch, R., Mervis, C., Gray, W., Johnson, D. Boyes-Braem, P. (1976). "Basic objects in natural categories." *Cognitive Psychology*, *8*, 382-439.

[Sha'ashua Ullman1988] Sha'ashua, A. Ullman, S. (1988). "Structural saliency: The detection of globally salient structures using a locally connected network." In *International conference on computer vision.*

[Shepard1981] Shepard, R. (1981). "Psychophysical complementarity." In M. Kobovy J. Pomerantz (editors, *Perceptual organization.* Hillsdale, NJ: Lawrence Erlbaum Associates.

[Tarr Bülthoff1995] Tarr, M. Bülthoff, H. (1995). "Is human object recognition better described by geon-structural-descriptions or by multiple-views? Comment on Biederman and Gerhardstein 1993." *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1494-1505.

[Todd etal1996] Todd, J., Koenderink, J., van Doorn, A. Kappers, A. (1996). "Effects of changing viewing conditions on the perceived structure of smoothly curved surfaces." *Journal of Experimental Psychology: Human Perception and Performance*, *22*(3), 695-706.

[Tsai1993] Tsai, F. (1993). "Robust affine invariant matching with applications to line

features." In *IEEE conference on computer vision and pattern recognition.*

[Tuller1967] Tuller, A. (1967). *A modern introduction to geometries.* Princeton, NJ: D. Van Nostrand Company, Inc.

[Ullman1979] Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, MA: MIT Press.

[Ullman1989] Ullman, S. (1989). "Aligning pictorial descriptions: An approach to object recognition." *Cognition*, *32*(3), 193-254.

[Van der Helm Leeuwenberg1996] Van der Helm, P. Leeuwenberg, E. (1996). "Goodness of visual regularities: a nontransformational approach." *Psychological Review*, *103*(3), 429-456.

[Van Gool etal1994] Van Gool, L., Moons, T., Pauwels, E. Wagemans, J. (1994). "Invariance from the Euclidean geometer's persepective." *Perception*, *23*, 547-561.

[Wagemans1999] Wagemans, J. (1999). "Toward a better approach to goodness: comment on Van der Helm and Leeuwenberg 1996." *Psychological Review*, *106*(3), 610-621.

[Wagemans etal1991] Wagemans, J., Van Gool, L. d'Ydewalle, G. (1991). "Detection of symmetry in tachistoscopically presented dot patterns: effects of multiple axes and skewing." *Perception and Psychophysics*, *50*(5), 413-427.

[Wagemans etal1992] Wagemans, J., Van Gool, L. d'Ydewalle, G. (1992). "Orientational effects and component processes in symmetry detection." *Quarterly Journal of Experimental Psychology*, *44A*, 475-508.

[Wagemans etal1993] Wagemans, J., Van Gool, V., L. Swinnen Van Horebeek, J. (1993). "Higher-order structure in regularity detection." *Vision Research*, *33*(8), 1067-1088.

[Weiss1993] Weiss, I. (1993). "Geometric invariants and object recognition." *International Journal of Computer Vision*, *10*(3), 207–231.

[Williams Julesz1992] Williams, D. Julesz, B. (1992). "Peceptual asymmetry in texture detection." *Proceedings of the National Academy of Science*, *89*, 6531-6534.

[Williams Jacobs1997a] Williams, L. Jacobs, D. (1997a). "Stochastic completion fields: A neural model of illusory contour shape and salience." *Neural Computation*, *9*, 837–858.

[Williams Jacobs1997b] Williams, L. Jacobs, D. (1997b). "Local parallel computation of stochastic completion fields." *Neural Computation*, *9*, 859–881.

[Witkin Tenenbaum1983] Witkin, A. Tenenbaum, J. (1983). "On the role of structure in vision." In J. Beck, B. Hope A. Rosenfeld (editors, *Human and machine vision.* New York, NY: Academic Press.

[Zerroug Nevatia1996] Zerroug, M. Nevatia, R. (1996). "Three-dimensional descriptions based on the analysis of the invariant and quasi-invariant properties of some curved-axis generalized cylinders." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*(3), 237-966.

[Zhu1999] Zhu, S. (1999). "Embedding Gestalt laws in the Markov random fields." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(11), 1170-1187.

Property	**Underlying Features**	**Minimal Viewpoint Invariant?**
Identity	two points, two lines	Yes
Proximity	two features	Minimal, quasi-invariant
Closure	curve	Yes
Trihedral vertex	three lines	Yes
Collinearity	three points, two lines	Yes
Smooth continuation	curve	Minimal, quasi-invariant
Parallelism	two lines	Yes
Convexity	three polygon vertices	Yes
	point on curve with curvature and figure/ground	Yes
Vertical/horizontal	one line	Only when camera must be upright
Right angle	two lines	When camera upright; picture may be rotated
Symmetry	4+ points or curves	Only when derived from midpoints between symmetric points

Table 1. The viewpoint invariant properties we discuss in this paper. Each property is formed by some underlying features. The properties of being vertical, horizontal, and right angles are viewpoint invariant only for viewpoints restricted to be upright. Symmetry is minimal only when evaluated in terms of the midpoints of symmetric points.