SRI International

CHOOSING A BASIS FOR PERCEPTUAL SPACE

Technical Note No. 315

January 3, 1984

By: Stephen T. Barnard, Senior Computer Scientist

Artificial Intelligence Center
Computer Science and Technology Division

SRI Project 5355

The work reported herein was supported by the Defense Advanced Research Projects Agency under Contract No. MDA903-83-C-0027.

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

Abstract

If it is possible to interpret an image as a projection of rectangular forms, there is a strong tendency for people to do so. In effect, a mathematical basis for a vector space appropriate to the world, rather than to the image, is selected. A computational solution to this problem is presented. It works by backprojecting image features into three-dimensional space, thereby generating (potentially) all possible interpretations, and by selecting those which are maximally orthogonal. In general, two solutions that correspond to perceptual reversals are found. The problem of choosing one of these is related to the knowledge of verticality. A measure of consistency of image features with a hypothetical solution is defined. In conclusion, the model supports an information-theortic interpretation of the Gestalt view of perception.

CONTENTS

FIGURES

1. Introduction

Why do we see the pattern of lines in Figure 1 as a right-angled corner? First, we must recognize that this is an **illusion**. (Let's call it the "right-angle illusion.") There is no strict, logical reason to interpret this figure in such a way: there are infinitely many three-dimensional spatial configurations of line segments that could have "explained" it. Nevertheless, we do see it in a special way — thus, we experience an illusion. Is it possible to understand this from a computational point-of-view?

The right-angle illusion does not depend on the three lines meeting at a common vertex. The pattern in Figure 2 evokes a comparable impression and it has no common vertex. The illusion is strengthened by rotating the pattern so that one line can be seen as vertical and the others as horizontal. This can be checked by rotating Figure 2 ninety degrees. Does the illusion still seem as vivid? On the other hand, more complex patterns, such as Figure 3, do not necessarily lead to right-angled interpretations, but, in these cases, the viewer is given additional constraining information beyond three line segments. If three line segments form two very acute angles, such as in Figure 4, they will not be seen as right-angled, but then no such interpretation is geometrically possible.

In summary, it seems that, in the absence of additional information, three non-colinear line segments will be seen as perpendicular lines in space, if such an interpretation is possible. There is strong experimental evidence for this hypothesis. Attneave and Frost [1] found that the perceived orientations of lines were highly predictable from hypothetical orientations implied by right-angled interpretations. Perkins [2] tested the ability to discriminate between right-angled and non-right-angled forms. He found that when an image could be explained by a right-angled interpretation it would almost always be perceived in that way.

At first, one might think the right-angle illusions too sparse to be meaningful. They contain so little information. Could they ever compare to real visual experience, with its abundance of data? Consider the familiar Ames-room illusion [3] (Figure 5). A weird, trapezoidal room is contrived to look normal (i.e., rectangular) from a particular point-of-view. Objects in the room are seen incorrectly: a man on one side of the room appears to be a midget, while another on the opposite side appears to be a giant. The Ames room illusion and the right-angle illusion have one critical point in common: in both cases, rectilinear perceptions are constructed from too little evidence. In the Ames room, furthermore, the effect is so strong that it dominates other important information for depth perception (such as size constancy). In an effort to make the distorted room look normal, our perception creates an incorrect geometrical interpretation that implies incorrect metric relations between objects. In effect, we perceive the *space* in which the room and the objects exist, rather than perceiving the room and the objects in isolation.

In this sense, the right-angle illusion is simply a minimal case of the Ames room.

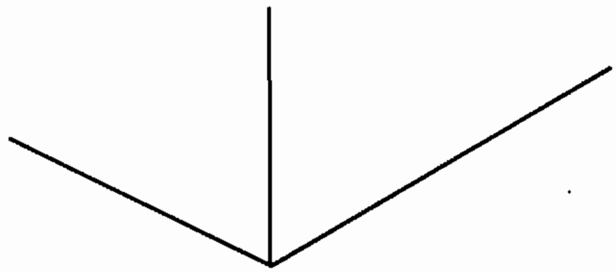


Figure 1: A Right-Angle Illusion

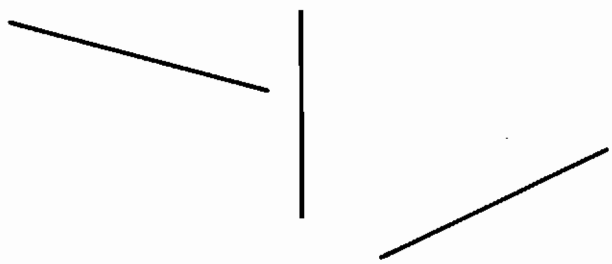


Figure 2: A Common Vertex is not Required

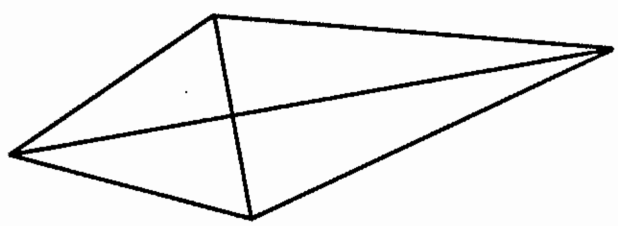


Figure 3: A Tetrahedron

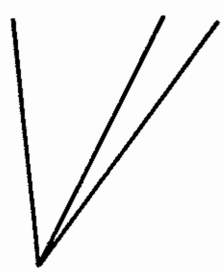


Figure 4: A Pattern that Does Not Admit a Right-Angled Interpretation

The difference between the two is that the Ames room contains abundant information for a rectilinear interpretation, whereas a pattern of three line segments contains the bare minimum. Any three mutually orthogonal lines from the Ames

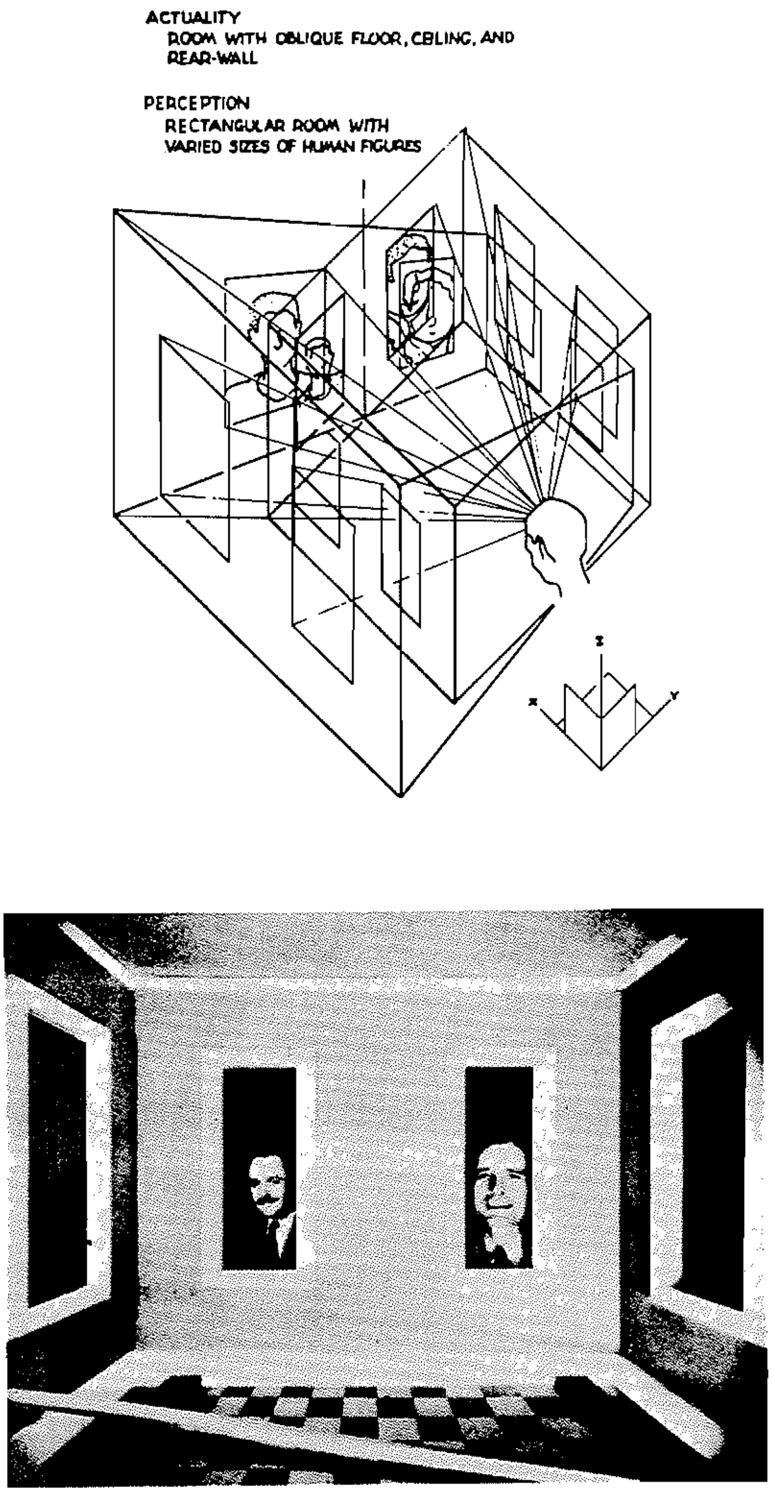


Figure 5: The Ames Room Illusion

room will produce an acceptable right-angle illusion, and, furthemore, all these illusions will be consistent in the sense of aligning with the natural coordinate system of the (imagined) room.

There are two reasons to be interested in this problem. First, a solution might suggest fundamental principles of perception. Human perception is an odd, complex, but remarkably consistent and efficient process. It "reasons" from incomplete evidence and almost never makes a serious error. Understanding this peculiar (and awesome) ability is the central task of vision research [4]. A second, pragmatic reason is that these patterns are quite common in images of natural scenes (see Figure 6). An algorithm that could make sense of them could contribute a basic capability to a larger machine vision system.

An early attack on a similar problem was directed at the so-called "blocks world." (See Mackworth [5] for a good summary of this line of research.) In a paper that pioneered the field of scene analysis, Roberts described a system for recognizing a small collection of simple, generic polyhedral shapes [6]. Whereas Roberts' methods produced complete metric descriptions of scenes, the blocks-world work that followed was aimed at segmentation and qualitative description. The methods were fundamentally syntactic and viewed the problem of blocks-world interpretation as a matter of parsing line drawings into allowable configurations of line and vertex types. The simplification of orthographic projection was introduced, and the effects of perspective were considered irrelevant. To the extent that metric constraints were used [7], [8], they were relatively weak and did not generalize in a straightforward way to perspective.

The approach presented here is quite different. A right-angle illusion, or a more complex image of an Ames room, or a blocks-world scene, or a natural scene such as Figure 6, imply certain interpretations *for geometrical reasons alone*. Specifically, intepretations that are in some sense "orthogonal" are preferred. A method for finding such interpretations for right-angle illusions will be presented. The approach is to seek a three-dimensional description that simultaneously accounts for the two-dimensional figure and the three-dimensional phenomenal perception. In contrast to the blocks-world results, the method is as easily stated for perspective as for orthography, and produces quantitative answers. It has a simple mathematical representation and computer implementation.

Figure 6: A Real Scene with Right-Angle Configurations

2. The Computational Model

In this section a formal mathematical model will be presented as a computational explanation of the right-angle illusion. The model consists of a method for constructing interpretations that are orthogonal and that are in the form of triplets of unit vectors. In essence, the interpretation constructs an alternative **basis** for the perceived space surrounding the viewer.

The best way to think of the method is as follows:

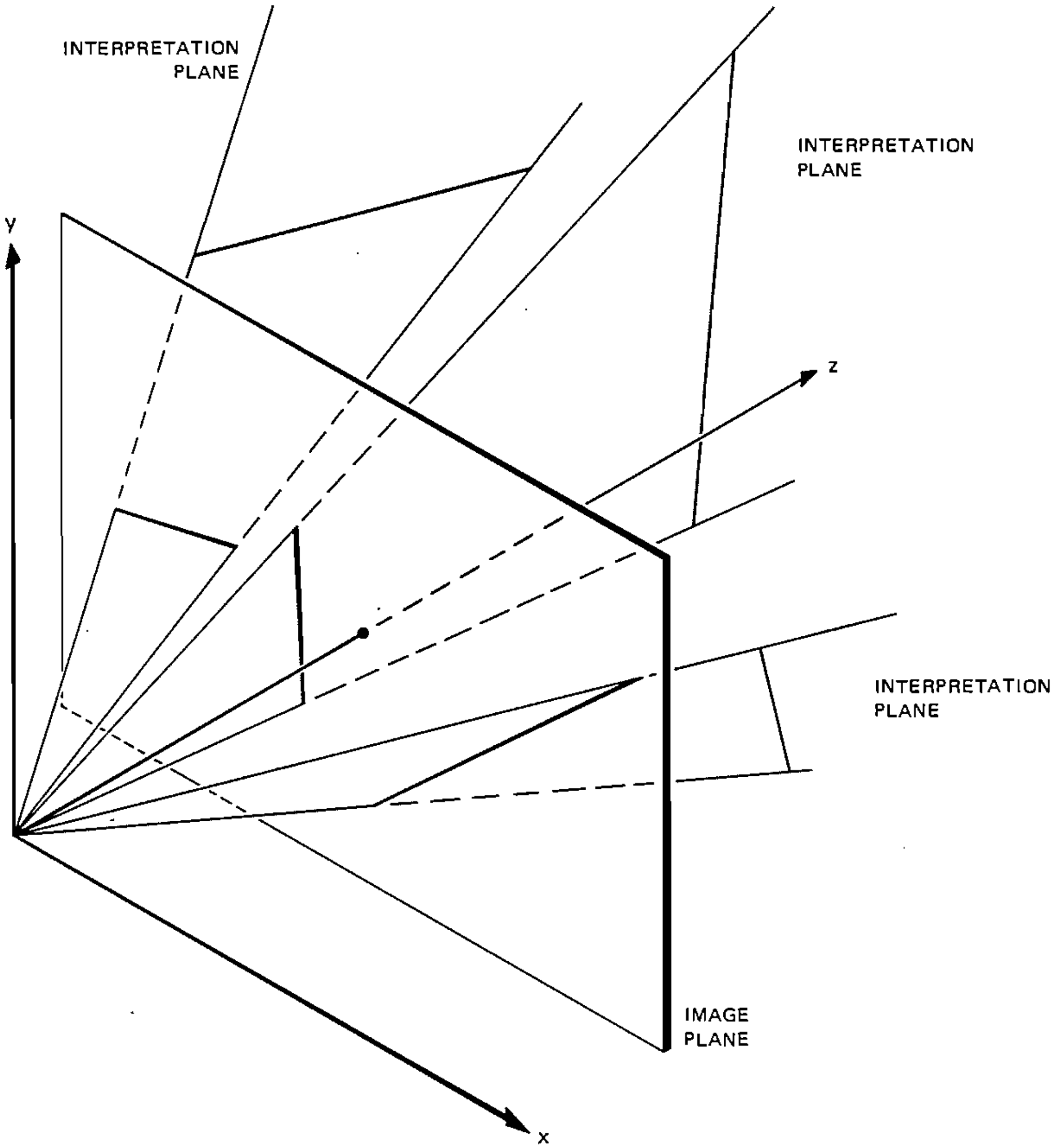
- A basis has six degrees of freedom (two degrees of freedom for each basis vector).
- Each line supplies one constraint. That is, each line constrains one of the basis vectors to a one-parameter family of vectors.
- The space of possible bases, therefore, is three-dimensional.
- The optimum basis is the one that is "most orthogonal." There will be two of these, in general.

2.1. The Most Orthogonal Basis

A system of three nonparallel, noncoplanar lines (orthogonal or not) defines a **basis** for a three-dimensional vector space. The goal is to find the basis that simultaneously is "most orthogonal" and is consistent with (i.e., explains) the two-dimensional pattern. This requires two elements: (1) a way to represent and generate the set of possible bases, and (2) a precise definition of the intuitive notion of "most orthogonal."

Let us call a pattern of three noncollinear, two-dimensional line segments (such as those shown in Figures 1 and 2) a **configuration**. We are not concerned with the length or the endpoints of the line segments. All collinear segments are considered identical. A configuration is assumed to be the result of a perspective projection of three lines in three-dimensional space (Figure 7). We will call any three such lines that produce a configuration an **admissible solution** to the configuration. Figure 7 illustrates how a configuration constrains the set of admissible solutions: the three lines of an admissible solution are constrained to lie in three planes determined by the line segments in the configuration. These planes are called **interpretation planes**. A configuration therefore can be characterized by the unit normals of three interpretation planes: (ϕ_1, ϕ_2, ϕ_3).

Clearly, the distances of the lines from the viewer are irrelevant. The lines are only required to have certain orientations and to lie in certain interpretation planes. We therefore consider all admissible solutions of a particular configuration consisting of lines of the same orientation to be equivalent. A class of equivalent admissible solutions defines an admissible basis. The basis consists of three unit vectors rooted at the origin (the center of projection), lying in the interpretation planes,



NOTE: Indicated lines in the image plane are the configuration; lines in the interpratation planes are an admissible solution.

Figure 7: Admissible Solutions

and parallel to the respective lines in the admissible solutions. These vectors can be generated in the standard, viewer-centered coordinate system. (This coordinate system is chosen such that the origin is at the projection point; the x, y, and z axes are in the directions right, up, and forward with respect to the observer; and the image plane is the plane $z = 1$. [1])

Let the three basis vectors be denoted by $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$. Remember that $\mathbf{e}_1$ lies in ϕ_1, etc. We can write a basis vector $\mathbf{e}$ in ϕ as a function of a scalar θ; for example, $\mathbf{e}_i(\theta)$ is in plane ϕ_i and at angle θ from the plane $z = 0$ (see Figure 8). The algebra for deriving this function is straightforward but somewhat tedious (see appendix).

We can now represent the set of admissible bases consistent with a configuration (ϕ_1, ϕ_2, ϕ_3):

$$S = \{[\mathbf{e}_1(\theta_1), \mathbf{e}_2(\theta_2), \mathbf{e}_3(\theta_3)] : -\pi < \theta_1, \theta_2, \theta_3 \leq \pi\}$$

Generating elements of this set is simply a matter of generating and substituting values for θ_1, θ_2, and θ_3.

The "orthogonality" of an admissible solution can be stated in a natural way as a triple product:

$$V = \mathbf{e}_1 \cdot (\mathbf{e}_2 \times \mathbf{e}_3)$$

This equation gives the volume of a parallelepiped associated with the three basis vectors (Figure 9). It is sometimes called the box product. The triple product has a maximum (or minimum) value of 1 (or -1) only when the vectors constitute an orthogonal basis. In the first case they form a left-handed basis, and in the second case a right-handed one. [2] The triple product has a value zero only when the three basis vectors are coplanar (i.e., linearly dependent).

We can find the most orthogonal basis by searching the three-dimensional space of admissible solutions for those with maximum or minimum V. In practice, there seem to be a unique minimum and maximum when an orthogonal solution is possible, and these extrema can be reached by the method of steepest ascent (or descent) from an arbitrary starting position. There is currently no proof of these conjectures, but they have held true for many different examples.

Figure 10 shows the starting point ($\theta_1 = \theta_2 = \theta_3 = 0$; i.e., the flat parallelepiped lying in the image plane), two intermediate solutions, and the final, optimal solution. The figures are produced by constructing parallelepipeds from the bases, centering them at the point $\langle 0, 0, 3\rangle$ in the viewer coordinate system, and projecting them into the image plane with hidden-line removal. The initial parallelepiped (shown in (a)) has zero volume because all the vectors lie in one plane. The next two (shown in (b) and (c)) have successively larger volumes (hence the associated bases are more orthogonal). The final parallelepiped (shown in (d)) is actually a cube, and its associated basis is truly orthogonal. Figure 11 puts the solution in context.

[1] Note that this is a left-handed coordinate system.

[2] Because we begin with a left-handed viewer coordinate system, we apply the left-hand rule when computing the cross product.

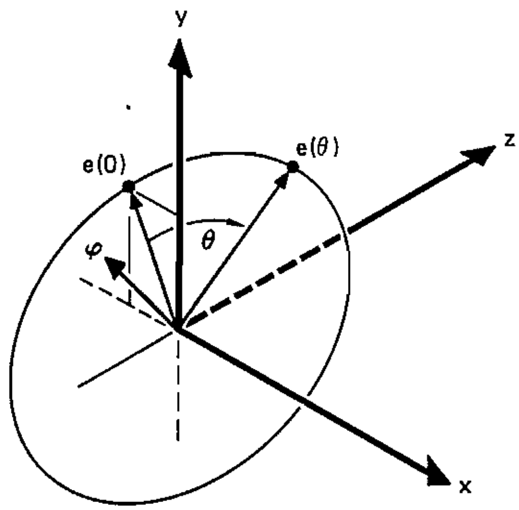


Figure 8: Basis Vector in an Interpretation Plane

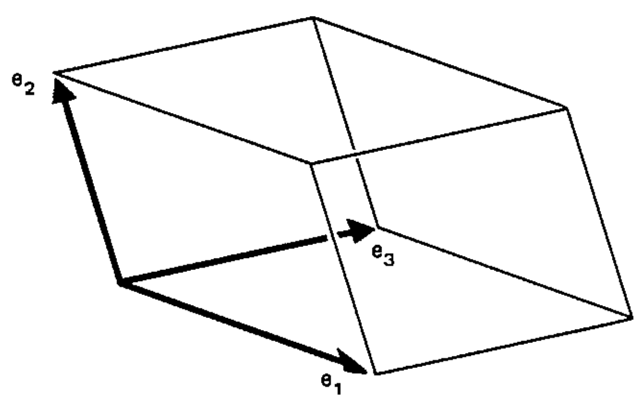


Figure 9: Parallelepiped Associated with Basis Vectors

At this point a discussion of the nature of the interpretation produced by this model is appropriate. It is a *scene-centered* (or object-centered) interpretation in the sense of orientation; that is, it decouples the natural orientation of the scene (or the object) from that of the viewer. It is a ***viewer-centered*** interpretation in the sense of position; that is, the origin of the most orthogonal basis is the same as the natural origin of the viewer (the center of projection).

2.2. Two Solutions: Which to Choose?

There are two ways to choose a most orthogonal basis: we can either maximize or minimize V. As mentioned above, a maximum V implies a left-handed basis, and a

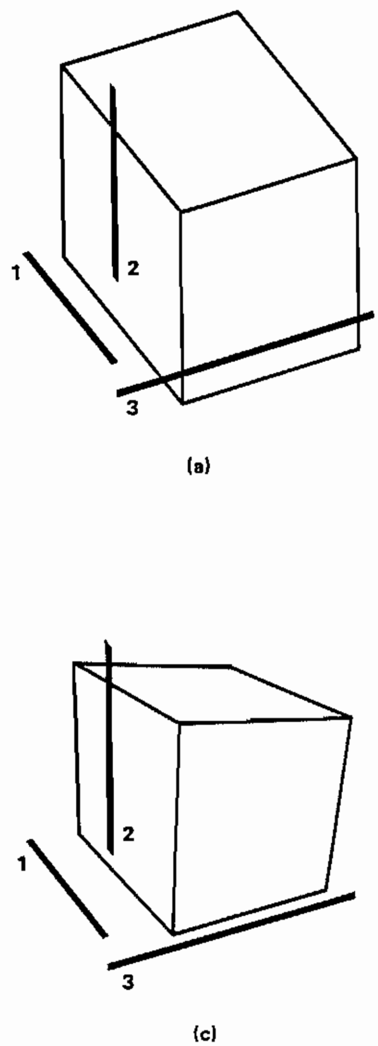


Figure 10: An Example

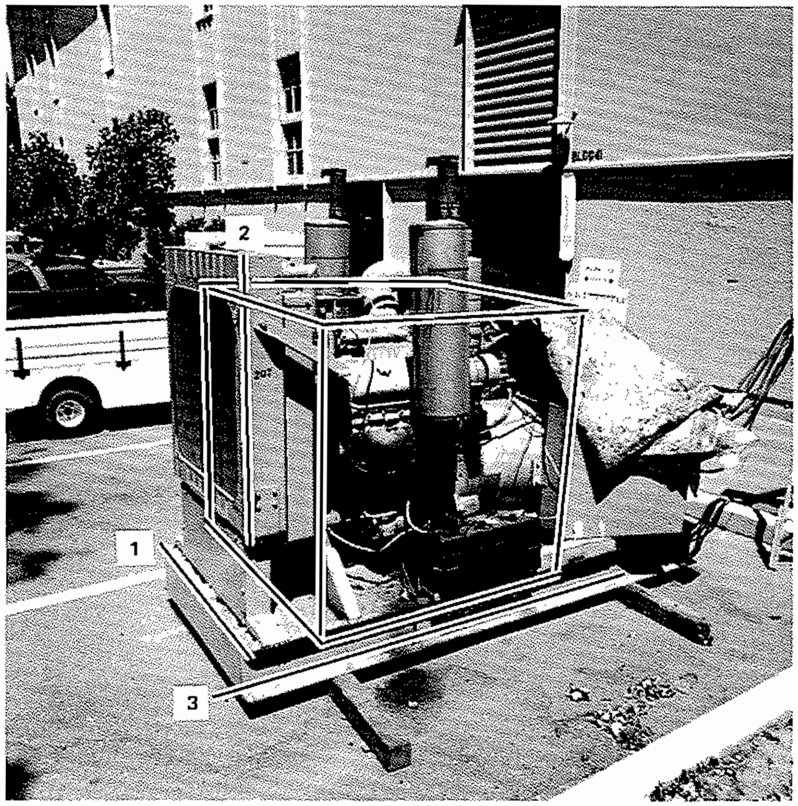


Figure 11: The Interpretation of Figure 10d in Context

minimum V a right-handed basis. The two solutions actually represent perceptual reversals. Is there any reason to choose one or the other? The handedness property is not significant because it is merely an artifact of the arbitrary direction sense of the interpretation-plane normals.

In Figure 12, a configuration and two alternate solutions are shown. For some reason, solution (b) does not seem to be as "good" as (a), even though, from a strictly mathematical point-of-view, it must be. The answer is suggested in our experiment of rotating Figure 2. The "good" solution is oriented in a way that is consistent with our notion of vertical and horizontal, but the other one is not.

In the everyday world, the effect of gravity imparts a special meaning to the "vertical" direction; similarly, while there is no unique "horizontal" direction, horizontal lines are constrained to be perpendicular to the vertical direction. All everyday scenes have a natural horizon, which may or may not be directly visible. Even if it is not directly visible, the horizon is fixed by knowledge of the vertical direction, which, presumably, is available from other visual cues and from the mechanism of the inner ear. When we view a picture, we prefer it to be aligned so that the natural horizon lies across the visual field normally (i.e., horizontally in the image).

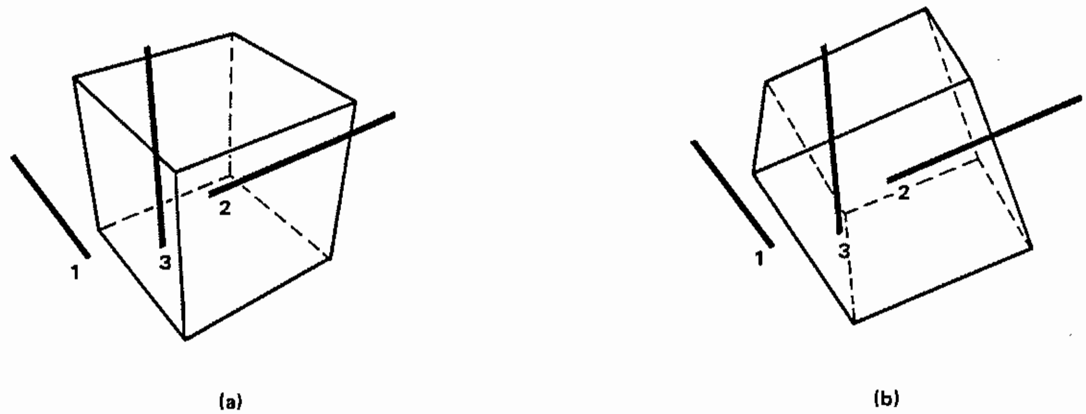


Figure 12: Two Alternate Solutions

It may be helpful to consider the relationship between the most orthogonal basis and the concept of vanishing points and lines. It is well-known that the perspective projections of parallel lines meet at common points in the image, called vanishing points. It has been shown that finding a vanishing point of a line is equivalent to finding the orientation of the line [9]. Hence, by finding the orientation of a basis vector, we determine the vanishing point of all lines parallel to it. A close examination of Figure 12 will show that, when the opposite edges of a side of the parallelepiped are extended, they intersect the extended lines of the configuration at vanishing points. Each of the two orthogonal bases therefore imply three vanishing points. Furthermore, if two of the vanishing points can be connected by a horizontal line (as in Figure 12(a)), the associated basis vectors can be interpreted as horizontal in 3-dimensional space.

2.3. Consistency

Suppose a line segment is added to a right-angle illusion. In Figure 13, three additional line segments, l_1, l_2, and l_3 are shown. Lines l_1 and l_2 seem to "fit" the rest of the illusion, while l_3 does not. That is, l_1 and l_2 can be interpreted as parallel or nearly parallel to at least one of the basis vectors, but l_3 cannot. In terms of vanishing points, we could consider all possible vanishing points of a line. If it had a possible vanishing point close to a vanishing point of a basis vector, it could be interpreted as parallel or nearly parallel to the basis vector.

Accordingly, given a basis $[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$ and a line segment l with possible interpretations $\mathbf{e}(\theta)$, we can state a estimate of l's consistency with the basis as:

$$C = \{\max_{\theta}(|\mathbf{e}(\theta) \cdot \mathbf{e}_i| : i = 1, 2, 3\} .$$

Each element of C is the absolute value of the cosine of the minimum possible angle between one of the basis vectors and a three-dimensional line that projects to l. If

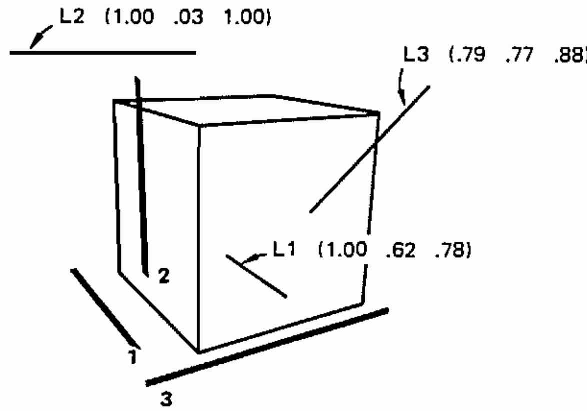


Figure 13: Consistencies of Three Lines

l can be interpreted as being parallel to a basis vector, the corresponding element of C will be one; otherwise, it will be less than one. The consistency values for l_1, l_2, and l_3 are shown in Figure 13. Note that l_1 is consistent with $\mathbf{e}_1$ only, but l_2 is consistent with both $\mathbf{e}_1$ and $\mathbf{e}_3$. Because l_2 is on the horizon, it intersects both of the horizontal vanishing points. Line l_3 is not very consistent with any basis vector.

This notion of consistency points to a way of finding the best interpretation for a large collection of lines, only some of which form a natural basis. One approach would be to select random triples of lines, solve for the most orthogonal basis, calculate the consistencies of the other lines, and choose the basis that was in some sense most consistent [10]. In the end, some lines would be inconsistent with the chosen basis and should be interpreted as having unknown orientations. A similar approach could be used to segment a scene into groups of lines that are related by virtue of being consistent with a particular basis.

It is quite possible for a configuration to lead to a most orthogonal basis that is not *actually* orthogonal (for example, three line segments separated by very acute angles, such as in Figure 4). In such a case, the method will yield a solution with $|V| < 1$. A nonorthogonal solution should probably be rejected. Of course, an orthogonal solution may also be incorrect (in the sense that it does not explain three lines in a scene correctly, because the lines are not really orthogonal). The important point is that, *given no more information than what is included in a single configuration*, the most orthogonal interpretation is reasonable.

3. Conclusions

The computational model presented in this paper is radically different from a widely prevailing view (e.g., see Marr [11]) that can be paraphrased as follows:

- Largely static, unintelligent processes convert an image into a collection of tokens, which comprise a discrete, explicit encoding of the information in the image (Marr's primal sketch).
- Evidence for local properties of surfaces (depth-from-viewer, orientation, curvature, reflectance, etc.) is extracted from the primal sketch by more-or-less autonomous processes (stereo, motion, shape-from-contour, shape-from-shading, etc.).
- This evidence is collected into a "2.5D" representation of the scene, meaning, an integrated description of surfaces in the coordinate system of the viewer.
- Instances of objects are found in the 2.5D representation (e.g., generalized cylinders).
- Finally, a description of the scene is constructed in terms of these objects.

In the model presented here, *there is no 2.5D sketch.* Furthermore, instead of a multiplicity of processes producing local, viewer-centered estimates, a single process produces a partial, scene-centered representation *directly.* The primal sketch retains its role, albeit in a more modest form — it essentially reduces to line-finding. This model is unconcerned with specific surfaces and objects. Instead, by producing a natural basis, it estimates a global property of the entire space surrounding the viewer.

This approach is most closely related to recent research on shape from contour [9], [12], [13], [14]. The general idea that relates this research is to backproject image contours onto planes of different orientations, and to choose, as the interpretation, the plane that simplifies some backprojected property. Several measures of simplicity are suggested. For example, Brady and Yuille use compactness, defined as the ratio of the area of the backprojected contour to the square of its perimeter; Barnard uses the uniformity of backprojected curvature; Witkin uses the degree of uniformity of the distribution of backprojected tangent directions. The model presented here yields interpretations not of the orientation of planes, but of space itself. Nevertheless, the philosophy is the same: to choose the most simple backprojection — in this case, simple in the sense of most orthogonal.

The Gestalt view of perception holds that percepts that are simple are preferred over those that are not. The modern version of Gestalt is that the percepts that can be most economically encoded are the ones preferred [15]. It is interesting that an orthogonal basis can be more economically encoded that a general one; that is, an orthogonal basis is **more redundant** than a general one. An orthogonal basis can be specified by any two of its basis vectors plus an indication of its handedness.

A general basis, however, requires a complete specification of all its basis vectors. The results in this paper are, therefore, consistent with, and lend support to, the information-theoretic version of Gestalt theory.

Of course, the model presented here is extremely simple and can in no way be considered a complete model of visual perception. Nevertheless, I feel that it does illustrate an important principle that is very likely to be used in human perception. Much work remains to be done to generalize and extend the model. The discussion of consistency in Section 2.3 points to one kind of generalization. The case of closed figures such as Figure 3 can be explained by an extension of the model, and this is a topic of current research. It remains to be seen whether the approach can be applied to curved contours and surfaces.

REFERENCES

1. Attneave, F. and Frost, R., The determination of perceived tridimensional orientation by minimum criteria, *Perception & Psychophysics*, vol. 6, no. 6B, December, 1969, 391-396.
2. Perkins, D., Visual discrimation between rectangular and nonrectangular parallelopipeds, *Perception & Psychophysics*, vol. 12, no. 5, 1972, 396-400.
3. Ittelson, W. H., **The Ames Demonstrations in Perception**, Hafner Publishing Company, New York, 1968.
4. Gregory, R. L., Will seeing machines have illusions?, in **Machine Intelligence 1**, N. L. Collins and D. Michie, Eds., American Elsevier, New York, 1967, 169-177.
5. Mackworth, A. K., How to see a simple world: an exegesis of some computer programs for scene analysis, in **Machine Intelligence 8**, N. L. Collins and D. Michie, Eds., American Elsevier, New York, 1977, 510-537.
6. Roberts, L.G., Machine perception of three-dimensional solids, in **Optical and Electro-Optical Information Processing**, J. T. Tippett et. al., Eds., M.I.T. Press, 1965.
7. Mackworth, A. K., Interpreting pictures of polyhedral scenes, *Artificial Intelligence* **4**, 1973, 121-137.
8. Kanade, T., A theory of origami world, *Artificial Intelligence* **13**, 1980, 279-311.
9. Barnard, S. T., Interpreting perspective images, *Artificial Intelligence* **21**, 1983, 435-462.
10. Fischler, M. A. and Bolles, R. C., Random-sample consensus: A paradigm for model-fitting with applications to image analysis and automated cartography, *Commun. ACM*, June 1981, 381-395.
11. Marr, D., **Vision**, W. H. Freeman, San Francisco, 1982.
12. Kanade, T., Recovery of the 3-D shape of an object from a single view, *Artificial Intelligence* **17**, 1981, 409-460.
13. Witkin, A. P., Recovering surface shape and orientation from texture, *Artificial Intelligence* **17**, 1981, 17-45.
14. Brady, M. and A. Yuille, An extremum principle for shape from contour, *Proceedings of IJCAI-83*, Karlsruhe, West Germany, August 8-12, 1983, 969-972.
15. Rock, I., **The Logic of Perception**, M.I.T. Press, Cambridge, Massachussetts, 1983.

A. Derivation of Functional Description of Basis Vectors

Given an interpretation plane represented by its unit normal

$$\phi = \langle \phi_x, \phi_y, \phi_z \rangle$$

we want to find an expression for the set of unit vectors in ϕ (Figure 8):

$$\{\mathbf{v} = \mathbf{v}(\theta) : -\pi < \theta \leq \pi\}.$$

We will develop and solve a system of three nonlinear equations in three unknowns: the components of $\mathbf{v}$.

The vector $\mathbf{v}(0)$ lies in the plane $z = 0$. We can impose an arbitrary directional sense to θ with

$$\mathbf{v}(0) \times \mathbf{v} = \phi \sin\theta. \tag{1}$$

This equation must hold, because $\mathbf{v}$ is perpendicular to the vector ϕ. (Refer to Figure 8. Remember that, because we use a left-handed coordinate system, we must apply the left-hand rule for a geometric interpretation of the vector cross product.)

Because $\mathbf{v}$ is a unit vector, it must satisfy

$$|\mathbf{v}| = 1. \tag{2}$$

Because $\mathbf{v}$ is in the interpretation plane ϕ, it must satisfy

$$\mathbf{v} \cdot \phi = 0. \tag{3}$$

Our system of equations is (1), (2), and (3). We will solve for $\mathbf{v}$ by first using (1) to get a simple expression for v_z, then substituting this in (2) and (3), eliminating v_y, and finally solving the resulting quadratic equation for v_x.

Let

$$D = \sqrt{\phi_x^2 + \phi_y^2}.$$

$\mathbf{v}(0)$ must be of the form:

$$\mathbf{v}(0) = (\frac{1}{D})\langle -\phi_y, \phi_x, 0 \rangle.$$

This is because it must simultaneously be in the direction

$$\phi \times \langle 0, 0, 1 \rangle = \langle -\phi_y, \phi_x, 0 \rangle$$

and satisfy

$$v_x^2 + v_y^2 + v_z^2 = 1.$$

Expanding (1), we get

$$\begin{aligned}\mathbf{v}(0) \times \mathbf{v} &= (\tfrac{1}{D})\langle -\phi_y, \phi_x, 0\rangle \times \langle v_x, v_y, v_z\rangle \\ &= (\tfrac{1}{D})\langle \phi_x v_z, \phi_y v_z, (-\phi_y v_y - \phi_z v_x)\rangle \\ &= \langle \phi_x \sin\theta, \phi_y \sin\theta, \phi_z \sin\theta\rangle.\end{aligned}$$

From the first component, we obtain our expression for v_z:

$$v_z = D\sin\theta. \tag{4}$$

Substituting (4) into (2) and (3) and expanding yields

$$v_x^2 + v_y^2 + D^2\sin^2\theta = 1 \tag{5}$$

and

$$v_x\phi_x + v_y\phi_y + D(\sin\theta)\phi_z = 0. \tag{6}$$

Solving (6) for v_y, substituting in (5), and collecting terms yields a quadratic in v_x:

$$D^2v_x^2 + 2\phi_z D(\sin\theta)v_x + D^2(\sin^2\theta)(\phi_y^2+\phi_z^2) - \phi_y^2 = 0. \tag{7}$$

We solve this for v_x:

$$v_x = (\frac{1}{D})[-\phi_x\phi_z(\sin\theta) \pm \sqrt{(\sin^2\theta)(\phi_x^2\phi_z^2 - D^2(\phi_y^2+\phi_z^2)) + \phi_y^2}]. \tag{8}$$

Now that we have expressions for v_x (8) and v_z (4), we can easily solve for v_y using (2).

Equation (7) has two solutions; the problem of which one to use can be resolved by observing that equation (3) is satisfied for two interpretation planes: ϕ and $-\phi$. This ambiguity results in the two solutions. Since the choice between ϕ and $-\phi$ is arbitrary, we can choose one and then use the appropriate form of (8).