Environmental Relations in Image Understanding: The Force of Gravity

John R. Kender

Computer Science Department
Columbia University
New York, NY 10027

January, 1983
Revised September, 1984

CUCS-47-83

1. Abstract

In this paper we show how assumptions and information concerning the external world properties of "horizontal" and "vertical" can aid in the analysis of images, even at the very lowest levels of processing. First, we review the pervasiveness of the force of gravity, and its influence on most natural image understanding systems. Next, we derive several fundamental mathematical results relating phenomena in both the gradient space and the image space to the external world attributes of horizontal and vertical. We then show how these results interrelate three imaging phenomena: the surfaces in the image, the external sensor parameters, and the environmental labels. We detail how, in general, specific information regarding any two of these phenomena can be used to quantitatively derive the third; occasionally one can do even better. Algorithms for such quantitative derivations are presented, including two based on the Hough transform. We further show how certain environmental perpendicularities can be exploited very efficiently, and even elegantly: ordinarily complex math simplifies to the extent that environmental distances can be directly read off the image. In this regard, they are analogous to the traditional line labellings of "concave" and "convex". The power of such environmental labels is then demonstrated by an analysis of the source of ambiguity in a simple illusion-like image configuration. The paper concludes with an analysis of the class of heuristics that have been invoked throughout. They are seen to be instantiations of the shape-from-texture meta-heuristics that "near implies preferred" and "preferred implies simple".

This research was sponsored in part by the Defense Advanced Research Projects Agency under contract N00039-82-C-0427.

2. Introduction

Many image environments are immersed in a force that strongly orients objects in a preferred way. The effects of this force are often so pervasive that environments which do not respond to it appear (and are often called) artificial. The very term "natural scene", vague though it may be, does at least seem to imply an image with just such a definite environmental orientation. Researchers would no sooner attempt to fully analyze such an image upside-down than they would if its colors had been permuted.

It is not difficult to be convinced of the influence that the presence of gravity has on the design of image understanding algorithms, especially in higher level processing. Often it is so strong that it deeply permeates the entire system as an implicit assumption. The assumption is made with good reason: higher level processing can be more efficient. Matching to models, for example, can start with both the detected object and the modeled object mutually aligned in the preferred (that is, the most probable) orientation.

However, we show in this paper that assuming the presence of gravity can aid the lower levels of image processing as well. This is a bit surprising, since many low-level routines do work--and ought to work--just as well with images inverted (or colors scrambled). Nevertheless, certain heuristics regarding the exploitation of "horizontal", "vertical", and other gravity-based environmental labels can make low-level shape recovery more efficient as well.

These heuristic assumptions, coupled with some fundamental mathematical results, can suggest methods and algorithms on the same level as other "shape from" methods, such as shape from shading or skewed symmetry [6, 10, 13, 20]. The heuristics themselves usually are based on the assumption of some preference: here, the preference for mutually perpendicular (horizontal and vertical) surfaces or lines. Thus, they can be seen as further members of the family of preference-based algorithms linked together by their derivation and use in a common methodological paradigm, called shape from texture [15].

These environmental labellings also have a family resemblance to the line labellings often used in the blocks world [8]. In an image of a trihedral scene, all straight lines are classified into one of three equivalence classes by a viewpoint-determined label (concave, convex, or occluding). Aside from reducing complex phenomena into simple and semantically suggestive symbols, they provide quantitative power, often permitting the exact determination of surface orientations. Environmental labels determine equivalence classes with similar semantic significance, and they are as quantitatively powerful. And unlike line labels, which fail when applied to perspective images [14], environmental labels are most powerful under perspective.

3. The Pervasiveness of Gravity

The presence of gravity introduces and maintains in "natural" environments a decided anisotropy. Its lines of force are parallel to each other in one specific, unchanging orientation. This orientation induces, usually by means of general energy minimization arguments, configurations that are themselves parallel: natural (as well as artificial) growth is often aligned with the field. Thus trees as well as buildings often have parallel sides, and are parallel to each other. Further, the "ground plane" is often actually planar, also in a minimizational reaction to the force: whether it is truly the ground, or artificially made so, as in a floor. The combination of these growth parallelism and ground planes further induce perpendicularities, again both natural and artificial. The junction of trees or animal legs to forest floor (or to their shadows), or the junctions of walls to ceilings (or object legs to floors), all occur in a limited class of orientations.

Natural systems that sense these environments have responded to these preferred orientations and alignments in direct and obvious ways. The eyes of many terrestrial animals lie on a horizontal line, as if to more adequately cover the horizontal surface that is the ground. (Many of the artificial aids to man's vision--television, the cinema-- also reflect this horizontal bias.) Few, if any, animals have a vertical or even random alignment. Animals with multiple eyes tend to have a bilateral symmetry which, together with their preferred whole-body orientation, induces a horizontal ocular arrangement. Even the flounder maintains the horizontal arrangement, at the cost of having to migrate an eye over the top of its head. To find examples of non-oriented eye placements, one has to investigate environments in which the gravitational force is negligible compared to other environmental forces. Thus, microscopic animals in aquatic, brownian-motion dominated worlds are free of gravity--and usually free of any preferred orientations in general.

Gravity influences human preferences and perceptions as well, sometimes in subtle ways. Artists know well the extent to which it alters awareness. One trick they use in order to more accurately render their drawings is to view the scene (and their work) upside-down, heads through their legs [17]. The perception of human faces in particular is gravity-sensitive; even a familiar face seen upside-down appears strange, with the forehead seeming bizarrely enlarged [4]. One of the many Ames illusions, the "star box" [11], shows the influence of gravity perhaps most starkly. In it, two vertically aligned points of light are shown to an observer in a darkened room. If the lights are both above the observer's viewing plane (the horizontal plane passing through both eyes), then the uppermost light appears closest. The reverse is true if the lights are below the viewing plane. In this most sterile of scenes, it is as if the points of light are interpreted as if they were attached to an imaged horizontal ceiling (or floor).

The effects of gravity are more subtle still. In our own field of computer vision, its influence is shown in work on the blocks world. Forgetting even the multitudinous mutual parallelisms and perpendicularities-to-ground that abound (often unexploited; however, see [1]), consider the way in which research results are presented and described. For example, in both Huffman's and Kanade's derivation of junction dictionaries from all possible three-space planar octants [9, 12], there is never any need to present the results with respect to any particular orientation. The work is gravity-free, yet most of the basic junctions are presented with at least one edge line vertically aligned on the page. Further, most if not all examples analyzed by their processing are aligned on a horizontal plane, even though the methods would work as well for, say, objects afloat in outer space.

Other examples of gravity's explicit and implicit involvement with image understanding can easily be given. In some domains, of course, it has no influence at all: for example, blood cell analysis. However, in most "natural" domains--or, equivalently, most "robotic" domains--its pervasiveness appears to be so extensive that a "natural" scene might very well be *defined* as one in which considerations of gravitationally induced orientations are non-negligible. In other words, a scene is a natural scene to the degree that it would be difficult to understand rotated or upside-down. (Thus, images of office interiors are about as natural as handwriting samples; both are more natural than most aerial photography; high magnification scanning electron micrographs are least natural of all.)

4. Basic Relations via Surfaces in the Gradient Space

Perhaps the first basic relationship that deals with the environmental labels "horizontal" and "vertical" are the terms used to define the degrees of freedom of the sensor itself. The sensor orientation terms "pan", "tilt", and "roll" imply a gravity-dependent coordinate system, and, in fact, are defined in environmental terms. Pan is sensor rotation in the horizontal plane; tilt is rotation in the vertical plane passing through the central visual ray. Roll is defined as rotation in the image plane, and its effect is therefore dependent on tilt and pan; in the abscence of roll, the image of an environmentally vertical plane that passes through the central visual ray is a retinally vertical line.

A second basic relationship is that, in terms of its use in computer vision, "horizontal" is simply a label for a unique, preferred surface orientation. In terms of the gradient space [18], it is a single labelled orientation point with coordinates $(p, q) = (p_h, q_h)$. Assuming that there is no roll in the sensor--that is, the y-axis of the image is the projection of an environmentally vertical plane--then this point simplifies to $(p,q) = (0, q_h)$.

This relationship is schematically depicted in Figure 4-1. The value of q_h is easily determinable: assuming the sensor is at a unit's distance from the ground plane and has no roll, then the central visual ray intersects the ground at q_h. Note that this value can also be obtained by a simple gravity sensor. The y-axis now lies in an environmentally vertical plane (Figure 4-2); further, due to the rotational coupling of the gradient space to the image space [18], the horizontal orientation has no p component.

It is not hard to show that every vertical surface must map into a gradient space point with coordinates $(p,q) = (p, -1/q_h)$. This fact follows from the general rule that the gradients of surfaces perpendicular to a given gradient (p_h, q_h) must satisfy the relation $pp_h + qq_h = -1$; every vertical surface is perpendicular to the horizontal. Thus, the one-dimensional family of vertical surfaces maps into the one-dimensional locus $q = -1/q_h$ (Figure 4-3) [16]. As a special case, if there is neither sensor roll nor tilt then the gradient space representation for the horizontal surface is infinitely far along the positive q axis, and the line of verticals becomes the p axis.

More generally, similar basic relationships hold even if there is a roll component. It is not hard to show that if there is information available about the sensor's tilt and roll, then the gradient space can be environmentally labelled by invoking the rotational coupling of the image space to the gradient space.

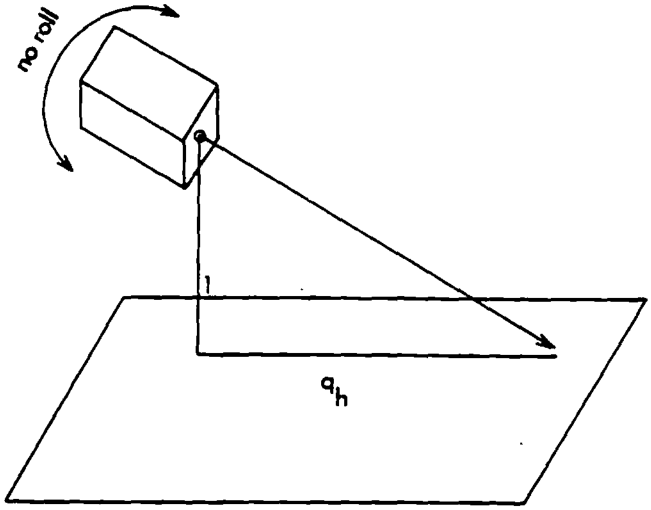

Figure 4-1: Basic relations: sensor configuration.

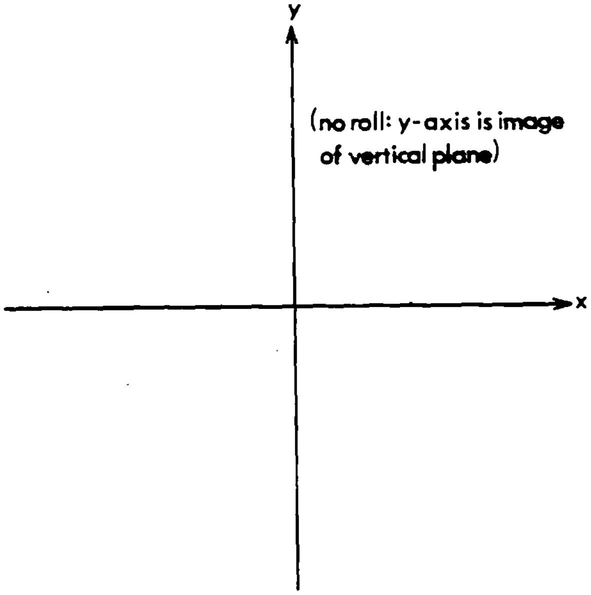

Figure 4-2: Basic relations: corresponding image space.

That is, if tilt is given as above by the angle whose tangent is q_h, and the roll component is given as Θ with respect to the unrolled sensor position, then the point in the gradient space corresponding to the horizontal is given by $(0, q_h)$ similarly rotated through Θ. The line of verticals rotates likewise: see Figure 4-4.

It is important to note that the above relations hold independently of any considerations of imaging projection. They are true for both orthography and perspective; they are true, in fact, even with no image at all. They describe the relations of the gradient space to environmental preference labels only. (Alternatively, one can use the analogous relations that prevail when surface orientations are recorded on

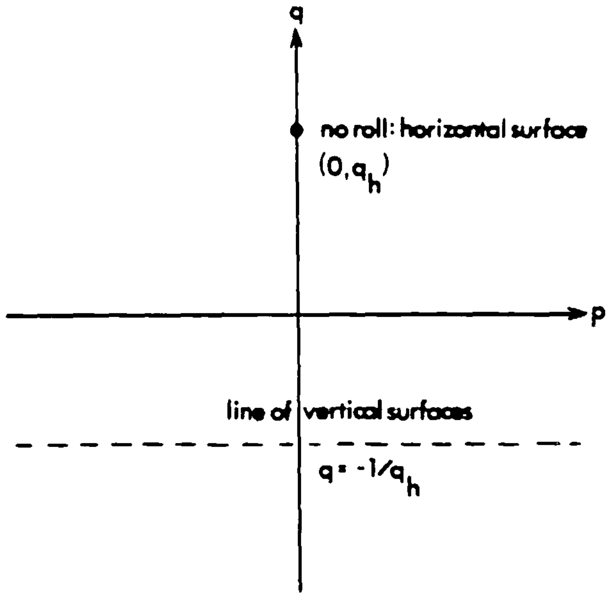

Figure 4-3: Basic relations: corresponding gradients.

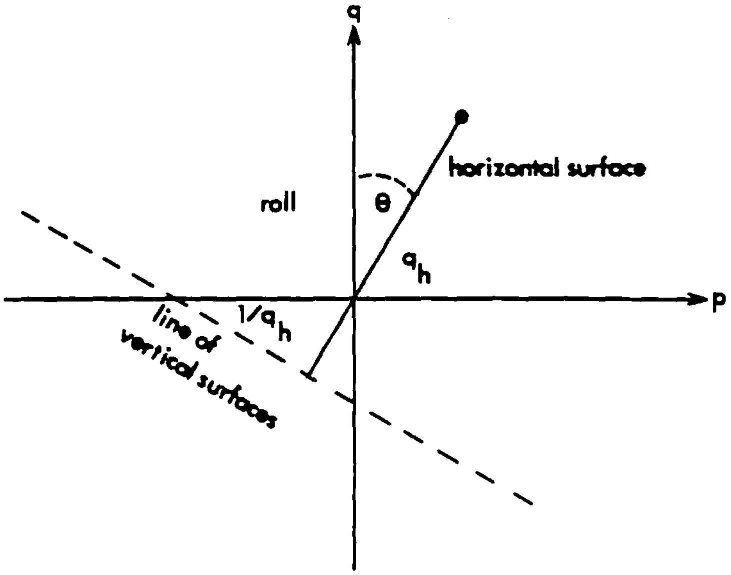

Figure 4-4: Basic relations: gradients under roll.

a Gaussian sphere map [7]). Further, they can be used in either direction: given sensor information, the gradient space can be labeled, and vice versa.

5. Basic Relations via Lines in the Image Space

At this point, we have not yet used any image information. In fact, in as much as surfaces exist in three-space, they cannot appear directly in an image at all. However, it is interesting to note that the same environmental labels of horizontal and vertical apply to lines as well, and to both lines in three-space and lines on the retina. Somewhat paradoxically, though, the size of the class of environmentally horizontal lines is one dimension greater than that of environmentally vertical ones; this is the reverse of the case with surfaces.

Environmental labels, environmental line segments, and the sensor parameters are related in several ways. To demonstrate them, consider first the case of perspective imaging where the sensor has no roll component. Scale the image plane in units of focal length; this will simplify the mathematics. Now image a scene consisting of vertical lines emerging from a horizontal plane: rather like a vast, stylized forest. The result is shown schematically in Figure 5-1.

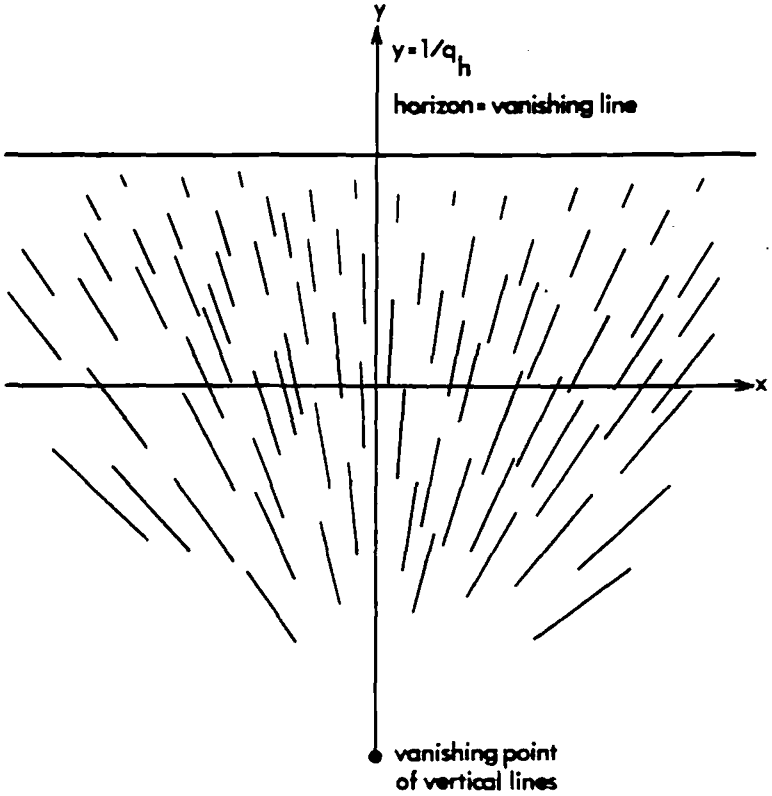

Figure 5-1: Vertical lines on a horizontal surface: perspective.

Because the class of environmentally horizontal lines is so large, they retain no distinguishing retinal features. That is, any line in the image can be the image of an environmentally horizontal line. About the only exploitable horizontal property is the horizon itself. This line, the limit of the projection of the horizontal plane, is a retinally horizontal. It has the equation $y = 1/q_h$. This follows from the basic relationship concerning vanishing lines: the plane with gradient (p, q) has vanishing line $px + qy = 1$; here $(p, q) = (0, q_h)$.

More interesting is the behavior of the environmental verticals. They form a more restricted class, and their images are more constrained. In particular, any environmentally vertical line must image into a retinal line that passes through the point $(x, y) = (0, -q_h)$. This follows as a special case of the analysis of vanishing points [18]. As the sensor's tilt increases so that its central visual ray approaches the vertical (i.e. as q_h approaches 0), this vanishing point of verticals approaches the image origin; simultaneously the horizon moves off in the positive y direction.

If the forest scene is imaged by an orthographic sensor, very little environmental information remains in the image (see 5-2). Nothing at all remains of the horizon. All environmentally vertical lines are imaged as retinally vertical lines; they have no finite vanishing point. Therefore, under orthography there are no image cues to sensor tilt.

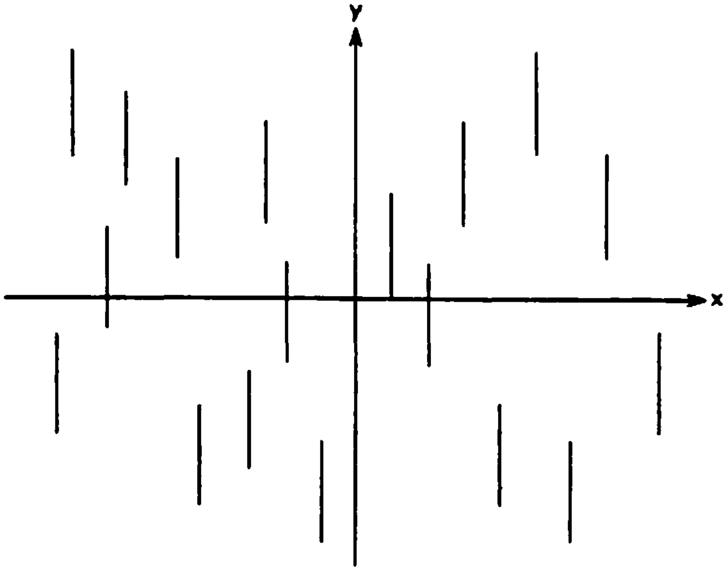

Figure 5-2: Vertical lines on a horizontal surface: orthography.

As with the gradient space relations, these image relations hold analogously under sensor roll. If the sensor is orthographic, then the parallel family of image verticals roll proportionately. If the sensor uses perspective, then the horizon and the vanishing point of verticals also roll proportionately and their expected locations are easy to compute, given tilt. The close relation between tilt and roll and the generated horizon is well known; it is exploited in the artificial horizon instruments of airplane cockpits.

Note that unlike the gradient space relations, however, these relations are not automatically reversible. That is, a given sensor configuration predicts definite image phenomena, but a given image phenomenon does not necessarily imply a sensor configuration. However, if the phenomenon can be environmentally labelled accurately (i.e. "horizon", "vertical vanishing point") then the implications about the sensor are correct. In general, though, this labelling must be done heuristically, as described below.

A summary of the basic relations is found in table 5-1

	Horizontal	Vertical
Surfaces	one-dimensional family $(p,q)=(p_h,q_h)$ specifies sensor	two-dimensional family $(p,q)=(p,-1/q_h)$ constrains sensor (1 degree)
Line in Scene	two-dimensional family unconstrained	one-dimensional family image through $(x,y)=(0,-q_h)$
Line on Retina	one-dimensional family	one-dimensional family

Table 5-1: Basic surface and line relations to labels.

6. Using the Gradient Space Relations

The relationships described above can be exploited in many ways. For example, given the sensor configuration, one can recover an environmentally labelled gradient space map as in Figure 4-3. If the sensor configuration is uncertain, the gradient space map (more simply, the gradient of a properly labeled horizontal surface) can be used to help calibrate tilt and roll.

It should be noted that the pan parameter can not recovered. In a sense, pan is "gravity invariant". That is, there is no information in an environmentally labeled gradient space map that would indicate pan. Pan does not even have any common environmental names. Perhaps the closest terms would be those used to describe compass directions: "north-by-northwest", etc. However, the magnetic force on which they are based seem to have negligible environmental influence; only a few natural systems are suspected of detecting it. Certain bacteria, for example, grow within themselves oriented grains of a magnetic iron compound. But even they use it not for directional discrimination, but rather as a guide to the vertical: they follow the natural magnetic flux lines (which are not truly horizontal) up to the ocean surface. One can speculate on how different visual perception would be if this planet's magnetic field were several orders of magnitude greater; as it is, this natural world does not seem to have a strong left-right preference. For example, although it is nearly impossible to find a newspaper photograph that has been printed upside-down, it is not unusual to find one that has been "flopped" left-for-right. Nevertheless, there may be environments in which it would be useful to augment a mobile robot's gravity sensor with a pan detector: a compass, or (for undersea work in stable environments) a prevailing current sensor.

Additional uses of the gradient space relations include the following. If the sensor parameters are known, then the determination that a given surface is horizontal uniquely specifies its gradient. The determination that it is vertical creates a linear constraint in the gradient space on which its gradient must lie. This constraint can be used with any other gradient space constraints: for example, those obtained by shape from shading, skewed symmetry, or shape from texture.

If the sensor parameters are unknown, then a determination that two non-parallel surfaces are vertical yields tilt and roll: their gradients generate the line of verticals in the gradient space. The determination of a single surface being vertical constrains tilt and roll to one degree of freedom; horizontal surfaces must be perpendicular to it.

6.1. A Hough-like Algorithm

Suppose we pose the more difficult problem in which there is *neither* a labelling nor sensor information. Nevertheless, both can still be (heuristically) recovered. Consider the additional assumption that all (or most) surfaces are either horizontal or vertical--an assumption often supportable in man-made environments. Then the gradient space representation (or the Gaussian map) of the surfaces in the scene can be analyzed for the presence of the characteristic point-of-horizontal/line-of-verticals configuration. This need not be an actual search for the line of verticals, although there may be some environments in which this is an efficient thing to do. Instead, it can be achieved using a type of Hough accumulator approach.

In broadest outline, this method has all existing surfaces vote for candidate horizontal surfaces. Once voting is done, the surface with the most votes is then presumed to be horizontal. Sensor tilt and roll, and the line of verticals are easily determined.

Voting is prescribed in the following way. Since a given surface is likely to be either horizontal or vertical, it votes once for itself since it may itself be horizontal. However, since it may also be vertical, it votes once for all surfaces perpendicular to it: in this case, at least one of these surfaces must be the horizontal. Graphically, this is displayed in Figure 6-1. A vote for the self is shown by a circle about the point; a vote for all perpendiculars is indicated by a dashed line. In the example, only one surface has received four votes; it is assumed to be horizontal.

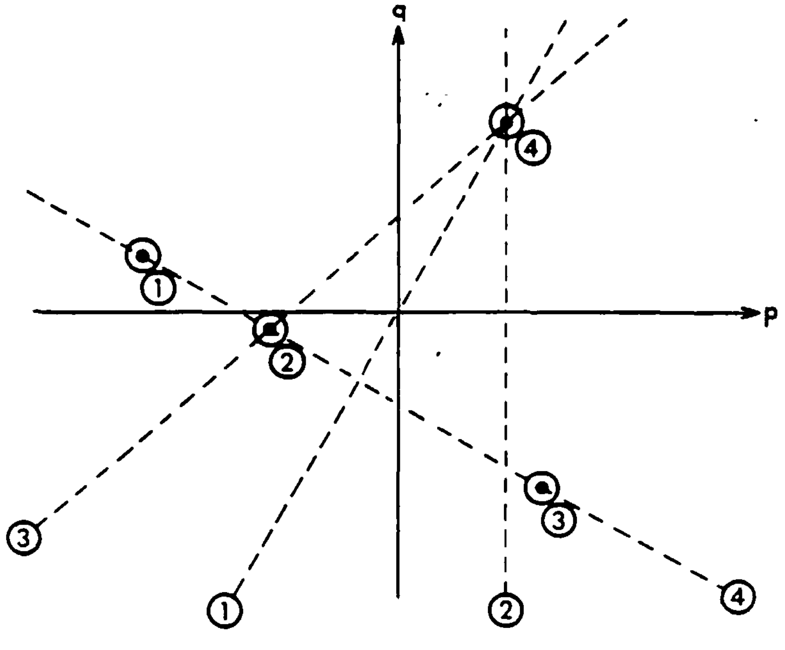

Figure 6-1: Hough scheme for finding ground planes.

6.2. A Critique of the Algorithm

This method is not without its problems, but it does have a virtue or two. The problems are manifest. Most critically, any such weighting scheme is heavily dependent on the given gradient space map, which in turn is affected by the environment and by the sensor position. Thus, if there is only one surface present, or even if there are two mutually perpendicular ones, there are no grounds by which to label anything horizontal. If there are two non-perpendicular surfaces present, the method considers them both vertical to a common horizontal (which does not appear in the gradient space.) If there are multiple surfaces, the voting is affected by the way in which the multiple surfaces have been recorded in the gradient space map: perhaps this map itself has been weighted. Lastly, the method is subject to the time and space problems that all Hough methods are plagued with: the space must be carefully quantized (a problem which is less severe on the Gaussian sphere), the line of votes must be calculated, votes must be distributed among accumulators proportionately, etc.

But the method does have some justifications. In particular, like most Hough transforms it can be made heuristically more efficient, and it is likely to be robust with respect to noise--which in this case are surfaces which are neither horizontal or vertical. Further, it works with surfaces that are curved verticals: building support columns, say, or drapery. In these cases, the gradient space map of the vertical surfaces is diffused along a line. Nevertheless the voting proceeds accurately, with each small quantum of the diffusion adding its small votes for its own perpendiculars. Perhaps most interesting is the result that the horizontal can be found even if there is no direct evidence for it in the gradient space: the ground can be "seen" even though it is "not there" (see Figure 6-2). This occurs when many environmentally vertical surfaces all vote for their perpendiculars; the one perpendicular they have in common must be horizontal, whether it is present in the gradient space map or not. (Some anecdotal evidence from the gravity-free environment of Skylab suggests that something similar may be at work with human beings. Astronauts tended to "lock in" to a subjective horizontal-vertical framework whenever a surface near their feet was within twenty degrees of their general body orientation. That is, the body itself was always viewed as "vertical"; external surfaces were labelled "horizontal" whenever they were close to the preferred perpendicular orientation.)

7. Using the Image Space Relations

The relations concerning image configurations can also be exploited in many ways. The simplest case is when all sensor information is known. One immediate result is that locations of both the horizon and the vanishing point of verticals are then also known, whether or not any phenomena suggesting them actually appear in the image. (If either *is* suggested by an image configuration, then that configuration can be assumed to be the proper one; its position can further calibrate sensor tilt and roll.)

Again, the more interesting algorithms occur when the sensor information is unknown, uncertain, or known only partially. Under perspective, if the focal point of the retina and the focal length of the sensor are known, then sensor tilt and roll can be found immediately from a line that has been correctly labelled

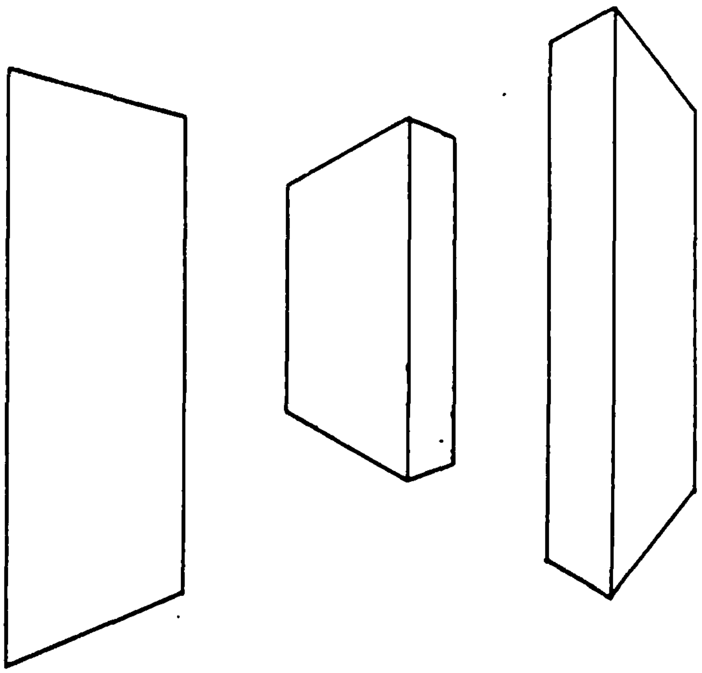

Figure 6-2: "Seeing" the ground plane.

as the horizon. The same is true if a pair of non-colinear lines are environmentally labelled as vertical: the intersection of their extensions give the vanishing point of verticals, and hence tilt and roll. (As with the gradient space, no image information provides pan.) If there is only one such vertical, then the tilt and roll are constrained to one degree of freedom; since the vanishing point of verticals can occur anywhere on this line, this constraint corresponds to the "vanishing gradient" of the line as defined in [18].

Still under perspective, if focal point, focal length, and roll are known, then only one environmental vertical line suffices to obtain tilt, as shown in Figure 7-1. In this case, the line that would be the image of a vertical plane through the focal point can be hallucinated; its intersection with the given vertical gives the vanishing point, which gives tilt.

Under orthography, neither the focal point nor focal length are required; they are "everywhere" and "infinity", respectively. But only roll is recoverable since the images of environmentally vertical lines no longer converge. However, roll can now be determined from a single image line that has been properly labelled as an environmental vertical.

Two questions remain: how are focal point, focal length, and roll obtained if they are unknown, and how are lines environmentally labeled?

Although complete sensor information is often available, occasionally--as in the case of an isolated photograph or a freely positioned sensor--some of it is not. Given that sense can usually be made of such environmentally detached images, it must be true that heuristics can be used to help quantify the missing

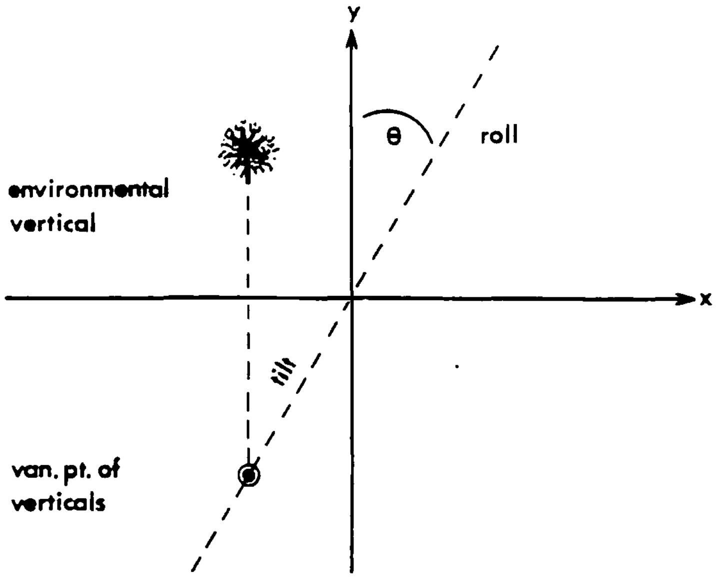

Figure 7-1: One environmental vertical gives sensor tilt.

parameters. There are many such means available, since the problem can be addressed at all processing levels of image understanding. For example, "ground truth" can help calibrate a sensor [2]. If one is dealing with man-made environments, one can use assumptions of multiple in-plane parallelisms and mutual inter-plane perpendicularities to obtain vanishing points; these constrain sensor location, sensor attitude, focal point, and focal length [15].

However, even at the lowest level of algorithms, fairly simple, purely environmental heuristics are possible. In the case of an isolated photograph, the focal point can be assumed to be the center of the photograph, and the focal length can be assumed to be a fixed ratio of the actual photograph dimensions. That is, the photograph can be assumed not to have been cropped.

The more pressing problem is with that of the free-floating sensor: the heuristic environmental labeling of lines for the determination of roll. Of course, given an uncropped image, one can always assume that was *no* roll, and that the image of the environmentally vertical plane through the focal point would appear as the image's vertical midline.

But one can perhaps do a bit better by first assuming that all retinally near-vertical lines, for suitable definitions of "near", are the images of environmentally vertical ones. Under perspective, a Hough-like scheme can then be used to help refine this heuristic labelling. Extend and weight all such lines, and take their most common intersection as the vertical vanishing point. Lines that do not pass through it lose the label. Tilt and roll come free. Under orthography, the method degenerates to one-dimensional histograming: since the images of true environmental verticals must be parallel, one only labels those lines that have the modal near-vertical (that is, the roll) orientation. These methods share all the usual properties of a Hough transform, good and bad. (The analogous methods for environmentally horizontal lines appear to be much weaker, given their unconstrained behavior in the image.)

A special case of such heuristic labeling of vertical lines is currently used in the 3-D Mosaic system [5]. It works on aerial views of buildings in Washington, D.C., taken with a sensor aligned directly downward. The vanishing point for verticals is therefore the image origin, and any line whose extension passes through the image origin should be heuristically labelled a building's vertical. This is exactly what the system does.

However, even if there is no indication of a probable near-vertical direction--the image has been circularly cropped, say--a related heuristic applies. It is that for many natural scenes taken from a mobile sensor, sensor tilt is often small or zero. In large part this is due to the fact that most environmental activity takes place on or near the horizontal plane through the sensor; in particular, navigation through a gravitational field is most concerned with goals, obstacles, or threats on about the same physical level above the ground plane as the sensor itself. The heuristic result is that the vanishing point of verticals is expected to be the most distant vanishing point of all, especially in environments without preferred pan orientations. This is evident in many cases of architectural rendering, where the most prevalent drawing technique is two-point perspective. Within it, environmentally vertical lines are drawn actually parallel, with the result that the most distant vanishing point is the infinitely distant vanishing point of the verticals. The heuristic is even supported by the oldest form of perspective drawing, Renaissance one-point perspective, in which verticals are also drawn actually parallel (although some horizontals are, too).

This heuristic is rather robust. To violate it, the sensor orientation must be unusual, in a way that can be quantified under some assumptions. Consider a scene in which there were two families of parallel lines, one of which is the vertical family. The heuristic incorrectly chooses the non-vertical family as the vertical only whenever the sensor is oriented more obliquely to the vertical than it is to the other family.

To simplify the analysis, assume the other family is a horizontal one. (The analysis for other cases, such as for multiple horizontal families or for families that are neither horizontal or vertical, is messier but similar, and the conclusions are nearly the same.) The vertical and horizontal families together create a tilt-pan environmental reference framework. Within it, tilt and pan together are zero when the sensor is aligned so that the image of both families are actually parallel (such as when the sensor is looking directly at a brick wall). In effect, these families establish a local compass frame; "north" is when pan is zero.

Deviations from this alignment cause tilt or pan or both to be non-zero, and the vanishing points of one or both families to become nearer to the image center. Most notably, under the common cases of pure pan (zero tilt), the vertical family remains imaged as parallel lines, whereas the horizontal family--or any other family--is almost always imaged with finite vanishing points. Under pure pan, then, the heuristic is exact.

Suppose now that tilt departs from the horizontal plane; the vertical family now has a finite vanishing point. The horizontal family can now be imaged only over a narrow range of pans so that its vanishing point remains more distant; the simplest case is a pan of zero. If tilt angles remain small, this range is approximately four times the tilt angle: the sensor can depart from the zero pan position plus or minus the tilt angle, and it can do the same in the exact opposite pan directions. (The exact solution involves solving a hyperbola very much like the Kanade hyperbola.) Therefore if tilt is, say, 10 degrees off horizontal, pan can be 10 degrees to the "east" or "west" of "north" or "south", without generating a vanishing point nearer than that of the verticals.

For the small tilts expected of a gravity-bound navigating sensor, then, the inaccuracy of the heuristic is approximately $2T/\Pi$, where T is the tilt in radians. If a record of past tilt distribution is available, the total expected inaccuracy can be found through integration; the more likely the sensor is horizontal, the less the inaccuracy. In fact, the heuristic fails completely (in the case of the other family being horizontal) if tilt regularly exceeds 45 degrees, since beyond this limit the vertical vanishing point is always the *nearest* vanishing point. It may be that the "dramatic" quality of photographs of buildings taken nearly vertically (from top or from base) comes in part from the failure of such a heuristic anticipation.

8. Exploiting Environmental Perpendicularities

The force of gravity also induces in the environment several types of perpendicularities. We have already described several algorithms that exploit the perpendicularities that are created upon horizontal surfaces by vertical surfaces or lines. We now show several ways in which to exploit the perpendicularity created upon horizontal *lines* by vertical lines. In the discussion that follows, we assume all sensor parameters are known. To simply the presentation, we further assume that there is no sensor roll, although nothing to follow depends on that fact.

In general, these algorithms are based on the observation that an environmentally horizontal line meets an environmentally vertical line at a right angle, and creates a vertical plane. Additionally, if the sensor parameters are known, the images of environmentally vertical lines can be hallucinated in abundance: they must only pass through the vanishing point of verticals (which is an ideal point in the case of orthography). Thus, all that is needed to define a vertical surface is the actual image of an environmentally horizontal line; see Figure 8-1.

The gradient of this hallucinated surface can be quantified. The knowledge that the plane is vertical already restricts its gradient to the line of vertical surfaces in the gradient space. But the fact that the environmental angle is a right angle itself generates another one-dimensional constraint in the gradient space. These two constraints can be intersected, usually resulting in a small, discrete number of gradients for the local vertical surface. Depending on the imaging geometry and the properties of the given horizontal, this gradient can be specified uniquely. In short, a line labelled as environmentally horizontal generates a well-specified vertical surface.

The second constraint, generated under the information that the image angle is environmentally right, is complex: it is a fourth degree curve in gradient space parameters p and q. However, under orthography,

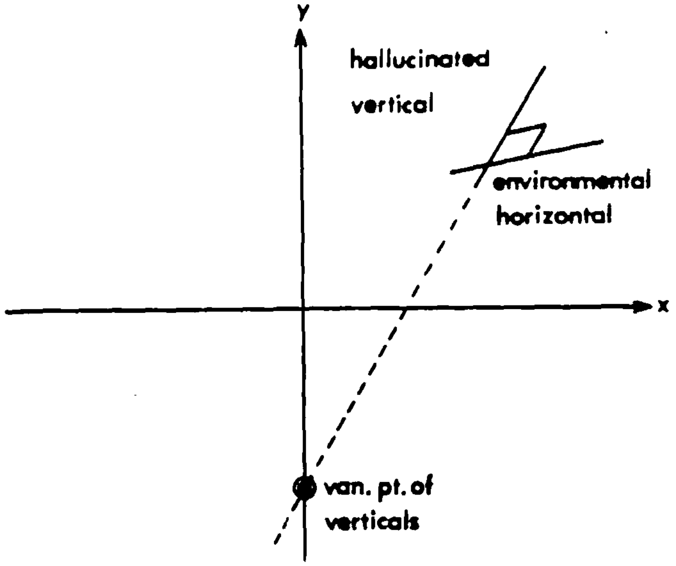

Figure 8-1: One environmental horizontal gives surface orientation.

or under perspective if the vertex of the angle lies on the focal point (local orthography), it simplifies to the Kanade hyperbola in p and q [15]. A further simplification occurs since one side of the angle is environmentally vertical: the angle can be drawn in both cases as in Figure 8-2. The vertex is at the image origin and the environmentally vertical side is aligned with the image of the vertical plane passing through the focal point. (With no roll, this image is the y-axis). One figure suffices for both the orthographic and special perspective cases because under orthography any image can be translated to the origin without affecting the gradient space.

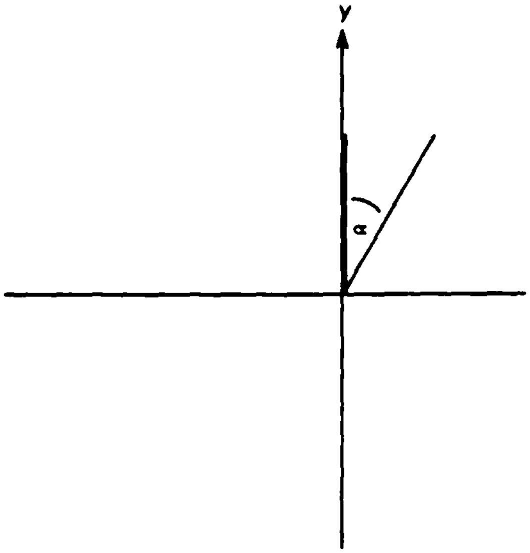

Figure 8-2: Simplest Kanade hyperbola. ımage.

The resultant constraint equation is still a hyperbola, but it is extremely simple: it is p = - cot (α) (q +

1/q), with p now a *one-to-one function* of q. As shown in Figure 8-3, this constraint is uniquely intercepted by the line of vertical surfaces, $q = -1/q_h$, for any value of q_h. Thus, under orthographic conditions the gradient of the generated vertical surface is uniquely defined. (Note that if the vertical line is an object edge and the horizontal line is the edge's shadow, then this gradient constrains the direction of the illuminant: see [19].)

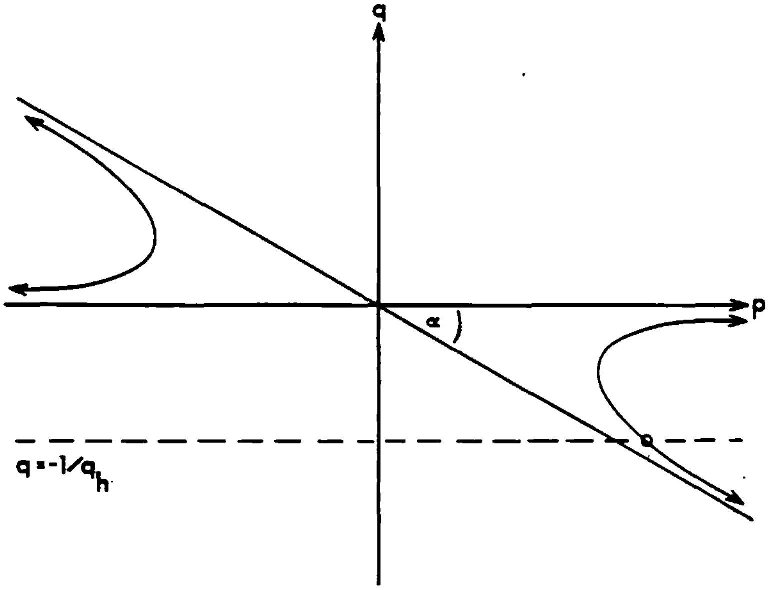

Figure 8-3: Simplest Kanade hyperbola: gradient space.

This special case hyperbola has several interesting properties. The first is that the minimum value of p always occurs at $q = -1$, *independent* of α. Since in orthographic photographs there is no indication of the sensor tilt, q_h, the observer is free to select a tilt at will. The choice of $q_h = -1$ guarantees that left-right slant is minimized (i.e. the surface "regresses to the frontal plane"). This value of q is equivalent to looking down at the horizontal plane at 45 degrees; this angle is commonly used in architectural drawing [17].

The second property is that under pure orthography all right angles with a vertical side behave identically, in one respect. Distances on the horizontal plane on which they stand can be read off *from the image*, independently of the angle α that their images form. Consider Figure 8-4. Let the vertex be at relative depth $z = 0$; distance increases towards the observer. Draw the retinally horizontal line $y = \cot(\alpha)$; the segment intercepted by the angle is of length 1. The total depth at the left intercept is $z = \cot(\alpha) / q_h$, since the vertical line has no p component and it increases in depth proportionally to the the sensor tilt. The total depth at the right intercept is the depth at the left *plus* the pure p component depth increase due to a movement of 1 image unit to the right. Thus, the depth at the right is $\cot(\alpha) / q_h - \cot(\alpha)(q_h + 1/q_h) = -\cot(\alpha)\, q_h$, using the function relating p to q_h.

This can all be summarized by stating that on any line $y = c$, the depth on the vertical leg is c/q_h and

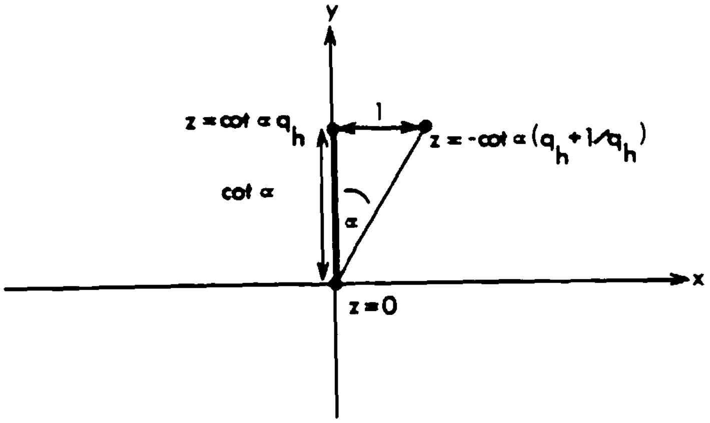

Figure 8-4: Right angle depths: image calculations.

the depth on the horizontal leg is cq_h, *independent* of α. In particular, any isolated right prism of whatever size, resting on a horizontal surface with known tilt, can be easily labelled for relative depth in the image itself, starting at any vertex and propagating depth changes outwards: see Figure 8-5. The prism need not be a parallelepiped; a triangular or hexagonal prism resting on its base (though not on a side) is parsed just as easily. Missing edges, as long as they do not separate the figure, are no problem, either. All that is required is that all edges be either vertical or horizontal. In this restricted blocks world, there is no need for even junction dictionaries, as long as sensor tilt is known. (It may even be possible to extend the method to handle occlusions, if "T" junctions are made to inhibit depth propagation to or from the leg of the "T".)

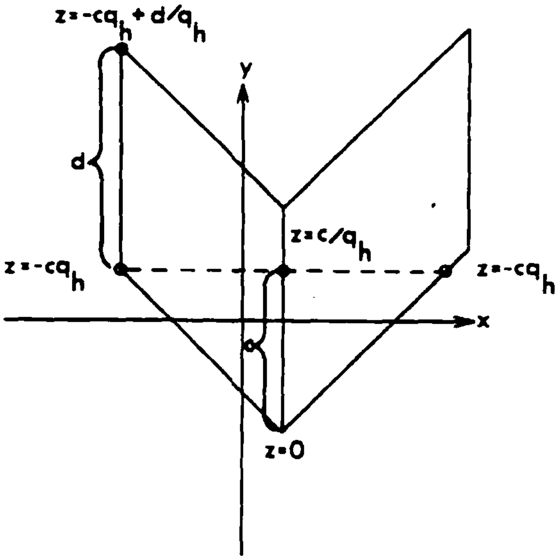

Figure 8-5: Right angle depths: propagation.

An alternate derivation of the relationships is shown in the side view of Figure 8-6. Along the plane of constant depth, at c units above the vertex, the environmental vertical has depth change c/q_h by similar triangles. Similarly, the horizontal's is cq_h. This side view also indicates the independence of depth calculations with respect to the image of the right angle; all that matters is the relative height in the image plane, and the environmental labels of vertical line or horizontal plane.

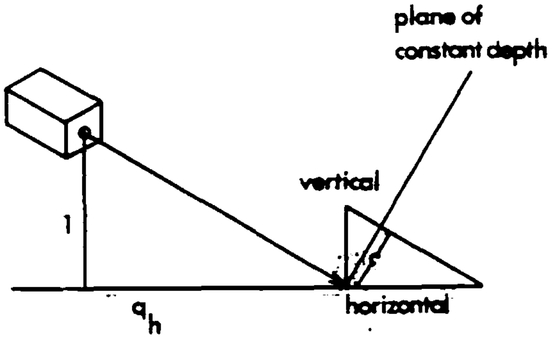

Figure 8-6: Right angle depths: side view.

9. Ambiguous Perpendicularities: The Importance of Line Labels

In the previous section, we gave algorithms for exploiting the perpendicularity that arises between horizontal and vertical lines. We demonstrate here that that configuration's power comes not from the perpendicularity per se, nor even from the fact that the surface that is formed is vertical, but from their individual environmental *line* labels. We show this by demonstrating that two *general* perpendicular lines, even within a environmentally vertical plane, give rise to ambiguous surface orientations. In this discussion, we make the simplest of assumptions: orthographic imaging with known sensor parameters and no roll; basically, this is a counter-example.

Consider Figure 9-1. It is the image of an environmentally vertical plane in which there is embedded a right angle. Neither side of the angle is environmentally horizontal or vertical, however. The constraint in the gradient space that the image generates from the assumption that it is environmentally right, is again the Kanade hyperbola: see Figure 9-2. However, because of the orientation of the image angle, this hyperbola is no longer a function. Further, some values of q have no corresponding p; that is, certain lines of vertical surfaces would not intersect this constraint curve. This is another way of saying that some values of sensor tilt q_h are incompatible with the interpretation of the image angle as a right angle in a vertical plane. (This was not so in the case with environmentally labelled side surfaces; here it is possible to create genuinely impossible scenes.) Worse, nearly every line of vertical surfaces that does intersect it intersects it twice. That is, for nearly every sensor tilt for which the image has an interpretation as a vertical plane, it has two possible gradients.

This example illustrates several points. First, it indicates the relative difficulties in computing (reflected in the relative difficulty of imagining) the properties of perpendicularities which not environmentally aligned. Developmental psychologists have noted this difficulty with young children's perception of the quadrilateral diamond shape: although it is a square rotated 45 degrees, and although squares are drawn fairly accurately, the diamond is most often drawn with unequal diagonals (usually vertically elongated). Secondly, the line labelling schemes for the blocks world have similarly proven to be more powerful than analogous surface labeling schemes [3], probably because lines constrain surface orientation and extent more severely than surfaces constrain lines.

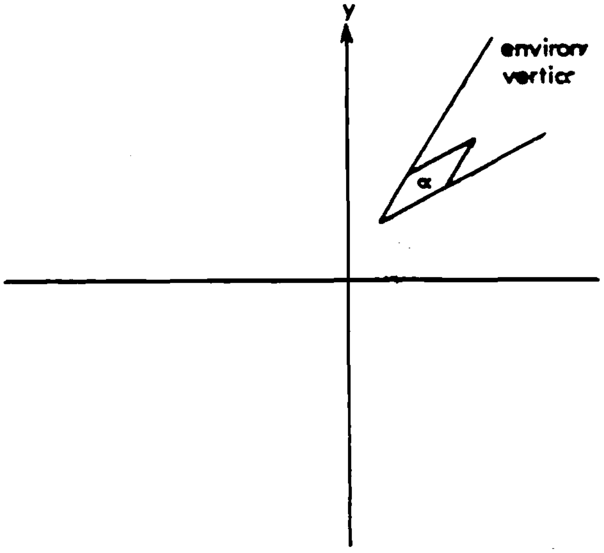

Figure 9-1: Ambiguous vertical planes: image.

It turns out that the two interpretations are somewhat difficult to visualize, probably because of cultural biases to see the world as perfectly carpentered in level and foursquare perpendicularity. Perhaps the best way to view them is with a physical construction, rather then by studying Figure 9-3. The following method seems to work. Insert a paper clip through the eraser end of a pencil. Trim the clip so that it has two equal legs, and bend them so that they form a right angle that lies in a plane perpendicular to the pencil itself. Now, using a sheet of graph paper as the horizontal ground plane, hold the pencil horizontally; the right angle formed by the clip is now in a vertical plane. View the clip with one eye from an angle fairly close to the vertical.

By rotating the pencil, both on its own axis and in the horizontal plane, search for the two configurations that present a retinal configuration like that of Figure 9-1. To make it easier to visualize, the retinal configuration can be drawn directly on the graph paper itself. Note that the value of p is related to the angle that the pencil makes with respect to the ruling of the graph paper. The case where the line of verticals is tangent to the hyperbola--the case of unique p--appears to correspond to the configuration where both legs are environmentally at 45 degrees to the environmental horizontal, although this is difficult to visualize or verify.

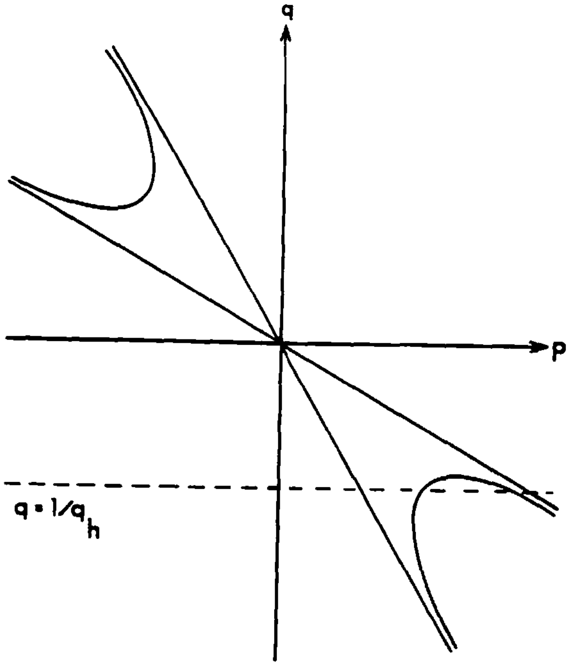

Figure 9-2: Ambiguous vertical planes: gradient space.

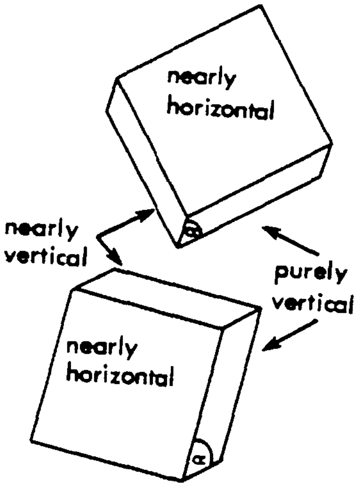

Figure 9-3: Ambiguous vertical planes: interpretations.

10. Discussion: The Shape-from-Texture Meta-heuristics

Although some of the relationships discussed in this paper have been absolute, many of them depended on heuristic assumptions. Most of the assumptions were of a similar form. The basic reasoning was as follows.

Certain preferred environmental objects create specific image configurations; for example, the images of environmentally vertical lines converge to a vanishing point. However, other environmental objects could also create the same configuration; for example, environmentally horizontal or oblique lines could also converge on the same vanishing point. The heuristics throughout assumed that the image configurations could be uniquely inverted as to cause: here, convergence implies environmentally vertical. More simply, the presence of an image feature similar to a preferred object's image features was taken as evidence for the preferred object. This "near implies preferred" meta-heuristic has proven useful in several other contexts, specifically shape from texture and skewed symmetry.

What sorts of environmental objects are preferred? One basis for preference is the simplicity with which image signatures can be inverted. For example, in the gradient space, both horizontal and vertical surfaces are easy to manipulate because their classes are small and well-defined; oblique surfaces are not. Horizontal and vertical surfaces are therefore preferred. This meta-heuristic that "preferred implies simple" has also proven useful in other contexts.

But perhaps the most evidence of the utility of the meta-heuristics is in the suggestion that foveation serves purposes other than an increase in resolution. Viewing perpendicularities off-axis under perspective leads to difficult mathematics; foveating them makes the math very simple. Thus, foveation helps to create simple signatures and helps define a preferred object. The implication for image understanding might be that all near-perpendicularities should be foveated; they might be the images of an easily determinable local vertical surfaces.

11. Acknowledgements

The author acknowledges stimulating discussions with Steve Shafer, and thanks Kerny McLaughlin for the artistic rendering of the figures.

References

[1] Barnard, S.T.
Choosing a Basis for Perceptual Space.
In *Proceedings of the IEEE Computer Society Workshop on Computer Vision: Representation and Control*, pages 225-230. April, 1984.

[2] Fischler, M.A., and Bolles, R.C.
Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography.
Communications of the ACM 24(6):381-395, June, 1981.

[3] Guzman, A.
Computer Recognition of Three Dimensional Objects in a Visual Scene.
Technical Report MAC-TR-59, Massachusetts Institute of Technology Artificial Intelligence Laboratory, 1968.

[4] Held, R.
Image, Object and Illusion.
W. H. Freeman Co., San Francisco, 1974.

[5] Herman, M., Kanade, T., and Kuroe, S.
Incremental Acquisition of a Three-Dimensional Scene Model from Images.
In Baumann, L. (editor), *Proceedings of the ARPA Image Understanding Workshop.* Science Applications Inc., September, 1982.

[6] Horn, B.K.P.
Understanding Image Intensities.
Artificial Intelligence 8(2):201-231, April, 1977.

[7] Horn, B.K.P.
Sequins and Quills--Representations for Surface Topography.
In Bajcsy, R. (editor), *Representation of Three-Dimensional Objects,* . Springer Verlag, 1982.

[8] Huffman, D.A.
Impossible Objects as Nonsense Sentences.
In *Machine Intelligence 6,* . Edinburgh University Press, 1971.

[9] Huffman, D.A.
Realizable Configuration of Lines in Pictures of Polyhedra.
In Elcock, E.W., and Michie, D. (editors), *Machine Intelligence 8,* pages 493-509. Edinburgh University Press, 1977.

[10] Ikeuchi, K. and Horn, B. K .P.
Numerical Shape from Shading and Occluding Boundaries.
Artificial Intelligence 17:141-184, 1981.

[11] Ittelson, W. H.
The Ames Demonstrations in Perception: A Guide to their Construction and Use.
Princeton University Press, 1952.

[12] Kanade, T.
A Theory of Origami World.
Artificial Intelligence 13(3):279-311, May, 1980.

[13] Kanade, T.
Recovery of the Three-Dimensional Shape of an Object from a Single View.
Artificial Intelligence 17(1-3):409-460, August, 1981.

[14] Kender, J.R.
Why Perspective is Difficult: How Two Algorithms Fail.
In *Proceedings of the National Conference on Artificial Intelligence*, pages 9-12. August, 1982.

[15] Kender, J.R.
AI Research Notes: Shape from Texture.
Pitman Publishing Ltd. , London, Accepted for publication, 1984.
Also available as Carnegie-Mellon University Computer Science Department Technical Report CMU-CS-81-102.

[16] Mackworth, A. K.
Interpreting Pictures of Polyhedral Scenes.
Artificial Intelligence 4(2):121-137, Summer, 1973.

[17] Morgan, S. W.
Architectural Drawing.
McGraw-Hill, New York, 1950.

[18] Shafer, S.A., Kanade, T., and Kender, J.R.
Gradient Space under Orthography and Perspective.
CVGIP 24(2):182-199, November, 1983.

[19] Shafer, S.A., and Kanade, T.
Using Shadows in Finding Surface Orientations.
CVGIP 22(1):145-176, April, 1983.

[20] Woodham, R.J.
Reflectance Map Techniques for Analyzing Defects in Metal Castings.
PhD thesis, Massachusetts Institute of Technology Artificial Intelligence Laboratory, June, 1978.
Available as AI-TR-457.

Table of Contents

List of Figures

List of Tables