

ENVIRONMENTAL LABELINGS IN LOW-LEVEL IMAGE UNDERSTANDING

John R. Kender

Department of Computer Science, Columbia University
New York, NY 10027

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. 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. 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".*

1 Introduction

Many image environments are immersed in a force that strongly orients objects in a preferred way. It is not difficult to be convinced that gravity influences the design of image understanding algorithms, especially in higher level processing. 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. Nevertheless, certain heuristics regarding the exploitation of "horizontal", "vertical", and other gravity-based environmental labels can make low-level shape recovery more efficient. These heuristics generate additional members of a family of preference-based algorithms for local surface orientation that are linked together by their derivation and use in a common methodological paradigm, called shape from texture [Kender 80].

2 Gradient Space Relations

Environmental labels first appear in the description

of the degrees of freedom of the imaging sensor. The terms "pan", "tilt", and "roll" imply a gravity-dependent coordinate system, and 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 absence of roll, the image of an environmentally vertical plane that passes through the central visual ray is a retinally vertical line.

A second class of environmental labels concerns surfaces. "Horizontal" is simply a label for a unique preferred surface orientation. In terms of the gradient space [Shafer 83], it is a unique point: (p_h, q_h) . Its coordinates can be obtained by a simple gravity sensor. If there is no sensor roll, this point simplifies to $(0, q_h)$. Further, every "vertical" surface must then map into a gradient space point of the form $(p, -1/q_h)$, since the gradients of surfaces perpendicular to (p_h, q_h) must satisfy the relation $pp_h + qq_h = -1$. Thus, the one-dimensional family of vertical surfaces maps into the one-dimensional locus $q = -1/q_h$. As a special case, if there is neither sensor roll nor tilt then the line of verticals becomes the p axis.

More generally, similar basic relationships hold even if there is a roll component. The gradient space can be environmentally labeled by invoking the rotational coupling of the image space to the gradient space. All these 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. Further, they can be used in either direction: given sensor information, the gradient space can be labeled, and vice versa.

It is interesting to note that the environmental surface labels of horizontal and vertical apply to environmental lines as well, and even to lines on the retina. However, 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.

3 Image Space Relations

Environmental surface labels, environmental line labels, and the sensor parameters are related in several ways. 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.

Unfortunately, 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 retinally horizontal. It has the equation

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

$y = 1/q_h$, as a plane with gradient (p, q) has vanishing line $px + qy = 1$.

More interesting is the behavior of the environmental verticals. 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. 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. Nothing at all remains of the horizon. All environmentally vertical lines are imaged as retinally vertical lines. Therefore, under orthography there are no environmental cues to sensor tilt.

All these image relations hold analogously under sensor roll. The close relation between tilt and roll and the horizon line is well known; it is exploited in the artificial horizon instruments of airplane cockpits.

Note that unlike the gradient space relations, these image relations are not automatically reversible. That is, a given image phenomenon does not necessarily imply any one sensor configuration. In general, environmental labeling must be done heuristically, as described below.

4 Using Gradient Space Relations

The relationships described above can be exploited in many ways. For example, given the sensor configuration, one can recover an environmentally labeled gradient space map. 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. (Pan can not be recovered; it is "gravity invariant". Pan does not even have any common environmental names: the closest terms for it are compass directions.)

Additional uses of these 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.

4.1 A Hough-like Algorithm

Suppose we pose the more difficult problem in which there is *neither* a labeling 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 can be analyzed for the presence of the characteristic point-of-horizontal/line-of-verticals configuration.

This can be achieved using a type of Hough accumulator approach. 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 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.

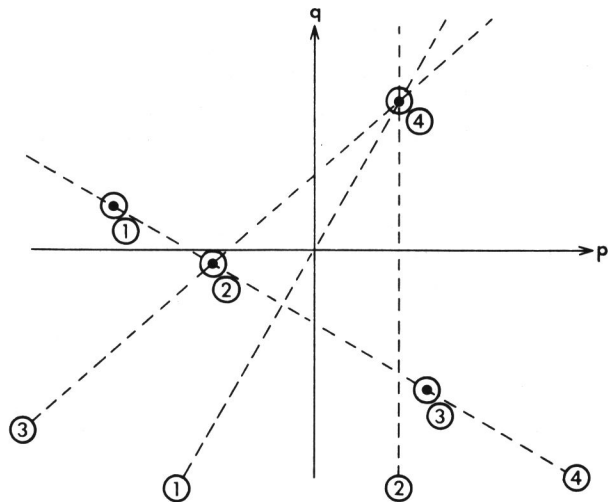


Figure 1: Hough scheme for finding ground planes.

4.2 A Critique of the Algorithm

This method is not without its problems. 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.

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. Perhaps most interesting is the result that the horizontal can be found even if there is no direct evidence for it in the image. This occurs when many environmentally vertical surfaces all vote for their perpendiculars; the one perpendicular they have in common must be horizontal.

5 Using 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 they actually appear in the image. (If either is suggested by an image configuration, then 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 labeled as the horizon. The same is true if a pair of non-colinear lines are environmentally labeled as vertical: the intersection of their extensions give the vanishing point of verticals, and hence tilt and roll. (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 vertical line, this constraint corresponds to the "vanishing gradient" of the line as defined in [Shafer 83].

Still under perspective, if focal point, focal length, and roll are known, then only one environmentally vertical line suffices to obtain tilt. 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 of verticals, which gives tilt.

Under orthography, 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 labeled as an environmental vertical.

Two questions remain: how is roll (and tilt) obtained if it is unknown, and how are lines environmentally labeled? Given that people can usually make sense of environmentally detached photographs, it must be true that heuristics can be used to help quantify the missing parameters.

One can assume 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 labeling. Extend (and weight) all such lines, and take their most common intersection as the vertical vanishing point; tilt and roll come free. Under orthography, the method degenerates to one-dimensional histogramming of image slope. (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 used in the Phoenix system [Herman 82]. 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. The method then becomes: any line whose extension passes through the image origin is heuristically labeled a building's vertical.

6 Using Linear Perpendicularities

We have 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 perpendicularities created upon horizontal *lines* by vertical lines. We assume all sensor parameters are known, and to simply the presentation, we further assume that there is no sensor roll.

These algorithms are based on the observation that an environmentally horizontal line meets an environmentally vertical line at a right angle, creating an environmentally vertical surface. If sensor parameters are known, the vertical line is redundant. It can be hallucinated, since it must pass through the vanishing

point of verticals (an ideal point, in the case of orthography). Thus, all that is needed to define a vertical surface is the image of an environmentally horizontal line.

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 generates a second one-dimensional constraint in the gradient space (often, the "Kanade hyperbola"). These two constraints can be intersected, usually specifying the gradient directly. Thus, a line labeled as environmentally horizontal often generates a completely specified vertical surface.

In the most general case, the second constraint is complex: it is a fourth degree curve in gradient space parameters p and q . However, under orthography, 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 . A further simplification occurs when one side of the angle is environmentally vertical: the angle can be drawn with the vertex at the image origin and the environmentally vertical side aligned with the y -axis. The resultant constraint is extremely simple: it is $p = -\cot(q+1/q)$, with p now a *function* of q .

This function 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 unique gradient is a constraint on the direction of the illuminant.)

This function has several other 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° ; this angle is commonly used in architectural drawing.

A 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. On any image line $y = c$, the depth on the vertical leg is c/q_h and the depth on the horizontal leg is cq_h , independent of α . In particular, any rectangular prism of whatever size, resting on a horizontal surface with known tilt, can be easily labeled for relative depth in the image itself. Starting at any vertex, depth changes are propagated outwards: proportional to $1/q_h$ on vertical members, and to q_h on horizontal.

7 Ambiguity in Labeling

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 the line's individual environmental labels. 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 the image of an environmentally vertical plane in which there is embedded a right angle, neither side of which is environmentally horizontal or vertical. 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 2. However, because of the orientation of the image angle, this hyperbola is no longer a function of either p or q . Further, some values of q have no corresponding p ; that is, 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 previous case of environmentally labeled sides.) 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.

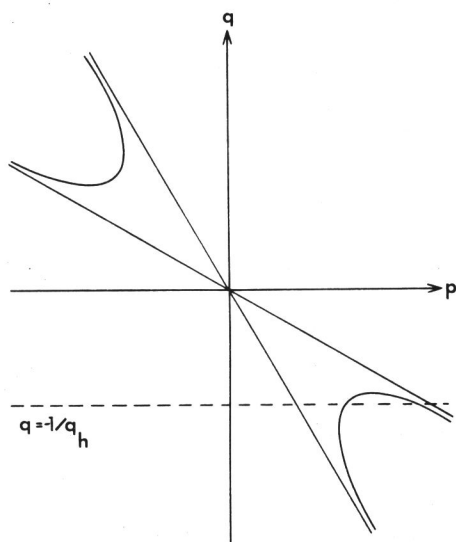


Figure 2: Constraint due to a vertical right angle.

It turns out that the two interpretations are somewhat difficult to visualize; see Figure 3.

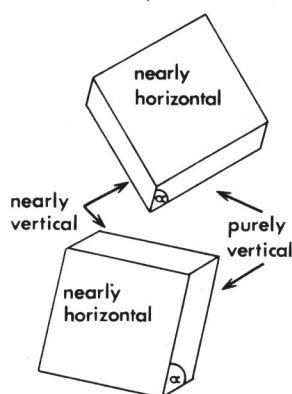


Figure 3: Ambiguity of right angle in vertical plane.

The case where the line of verticals is tangent to the hyperbola probably corresponds to the configuration where both legs of the right angle are at 45° to the environmental horizontal.

8 Discussion

Many of the relationships discussed in this paper have depended on heuristic assumptions 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 vanishing point. The heuristics throughout assumed that the image configurations could be uniquely inverted as to cause: here, convergence would be taken to imply 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.

Acknowledgements

I thank Kerny Calaway for her graphic skills.

References

- [1] 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.
- [2] Kender, J.R.
Shape from Texture.
PhD thesis, Carnegie-Mellon University, Computer Science Department, November, 1980.
- [3] Shafer, S.A., Kanade, T., and Kender, J.R.
Gradient Space under Orthography and Perspective.
Computer Graphics and Image Processing, To appear, 1983.