What Is A "Degenerate" View?

John R. Kender*
David G. Freudenstein

Department of Computer Science
Columbia University
New York, NY 10027-6699

1 Abstract

In this paper**, we attempt to quantify what is meant by the terms "degenerate view'*, and its relatives, "characteristic view", "visual event", and "general viewing position". We propose that the definition of degeneracy is itself degenerate, taking on differing meanings at different times. We claim (at least for the case of polyhedra) that one can only speak of a two-dimensional stimulus as being degenerate *with respect to* a given heuristic for inverting the image function. Additionally, we show that given the finite viewing resolution of a two-dimensional retina, in practice the concept of a characteristic view is often not characteristic of real imagery. Even precisely defined general viewing positions are sensitive to camera acuity: any viewpoint ceases to be characteristic at some resolution, and non-characteristic views are not vanishingly improbable. We provide initial quantitative estimates on these probabilities for some simple cases, and relate them to a minimal disambiguation distance. It follows that an aspect graph is less a discrete graph, and more properly a partitioning of the surface of the viewing sphere into "fuzzy" regions of non-zero area: an aspect *map.* This viewpoint is more in keeping with recent and proposed work on optimal viewing strategies.

2 Introduction

Robotic vision systems must both obtain images and analyze them. However, a primary characteristic of many realistic imaging situations is that the data acquisition is much less costly than the subsequent data analysis. In such domains it is therefore reasonable to dedicate significant computational effort towards the task of calculating an optimal viewing point for the next image capture. Defining and obtaining this optimum is necessarily probabilistic; it must incorporate an understanding of the limits of resolution of the camera, and of the limits of resolution of the placing agent The overall goal is to obtain maximal information from a sequence of inexact images in inexact placements, while minimizing some work function which expresses the relative costs of image acquisition and image analysis. Such calculations necessarily place a heavy premium on avoiding what are often referred to as "degenerate views".

Nevertheless, it is not apparent what makes a view degenerate, how such a view is recognized or forecast, or even whether such views are rare or commonplace. Thus, the first concern of this paper is to define and quantify the meaning of the term "degenerate", and to show the varying imaging contexts in which it can arise. Secondly, we suggest a representation useful for calculating the likelihood of such views (whatever their definition); it takes the form of mappings over the viewing sphere. This representation extends existing work on aspect graphs by explicitly incorporating the known limits of visual acuity. It also leads directly to methods of associating with each view a probability of its being attained, and the placement cost of attaining such a view.

3 Degenerate Viewpoints in Theory

If it is to be useful, a representation for the views of a three dimensional object or object assembly must give some insight into those viewing positions which are less helpful in resolving ambiguities of object structure, position, or orientation. We present two common views of what such degeneracy is, show that they are deficient, and redefine them in ways that are more quantifiable.

3.1 Slight Movements Giving Drastic Changes?

Perhaps the simplest example of ambiguity is the case of a head-on view of a cube, which is ideally imaged as a square. Such an image has often been noted as giving no information as to the three-dimensionality of the object, and has therefore been described as a degenerate image (see for example, [Kanade 80; Sabbah 82]). Degeneracy in this context, however, refers to the fact that a "slight" change in the viewpoint which generated the image would cause a "drastic" change in the image (sometimes called an "image event").

This definition (and related descriptions of what makes a viewing position general or an image characteristic [Chakravarty 82]) is inadequate in two ways:

1. It is vague with respect to the meanings of "slight" and "drastic".
2. It does not encompass all the phenomena that would seem to be properly described as examples of degeneracy.

What changes drastically in the cube-as-square image can be characterized in many ways: the number of regions change, the topology of the image regions is altered, apparent symmetries are modified, and (if lines are labeled in the Huffman-Clowes manner) the junctions are relabelled (cf. [Lavin 74; Thorpe and Shafer 83]). Still other derived properties of the image change, too. More generally, depending on the means of analysis, this view of the cube would be called degenerate if a slight change in viewpoint would alter the "quality" of the ensemble of extracted image features used in shape analysis. A rigorous definition of this "quality" change must then include the requirement that a qualitative change is one that ultimately affects the derivation of those object's "semantic" properties (such as identity, scale, rotation, coloring, etc.) considered important by the system.

The drastic change is therefore a drastic change in *interpretation,* not in image. Therefore the perception of a drastic change can vary from system to system. For example, if the system distinguishes cubes from spheres by detecting the presence or absence of long straight lines, the cube-as-square cannot be

*Thii research was supported in part by ARPA grant #N00039-84-C-0165, by a NSF Presidential Young Investigator Award, and by Faculty Development Awards from AT&T, Ford Motor Co., and Digital Equipment Corporation.

** A more detailed version of this paper, Including a survey of multiple viewpoint representations, as well as complete references and appendices, may be found In [Render and Freudensteln 87] considered degenerate. (And, in net, if the cube is the only model in the system at all, no view is ever degenerate.)

What is hiding behind mis implicit definition of "drastic" is the interpretation equivalence relation; a drastic change is a change of interpretation equivalence classes. However, since in many systems the interpretation classes are inherited from the feature equivalence classes, commonly the drastic change has been attributed to image characteristics alone.

Similarly, the notion of "slight movement" is imprecise. What is usually implicit in such definition is that there is at least one direction in which an arbitrarily small movement of the camera causes the drastic change. (Usually the direction is on a line perpendicular to an image edge). The meaning of "arbitrarily small" only appears to make sense when taken in the sense of mathematical analysis. That is, the drastic change must occur for positive movements of magnitude less than some epsilon, in the direction of degeneracy.

Such a definition would imply that degeneracy can be qualified as a matter of varying degree, although this is apparently never stated. That is, what can vary from degeneracy to degeneracy is the number of possible ("qualitative") drastic changes, and the relative number of directions in which such resolutions (or non-resolutions) occur. For example, the cube-as-square image can resolve itself into an image with either two or three regions, with die two region image possible only for four discrete directions of camera movement All other directions resolve it into an image with three regions, even if some of those regions are vanishingly thin. Further, some "degenerate" views sometimes do not resolve at all. For example, the cube imaged as two rectangles remains-two rectangles for two discrete directions of movement, resolving itself into three regions under movement in all other directions. The space of allowable degeneracies is apparently very large, and perhaps can be quantified in absolute terms as some measure defined on the ways in which the view fails to resolve into something more "characteristic".

The converse of the common "small gives drastic" definition is perhaps easier to implement This converse definition it stated as follows: An image is seen from a "general viewpoint" if there is some positive epsilon for which camera movements in any direction can be taken without effect on resulting semantic analysis. Here, too, the definition is based ultimately on system performance; an image can be degenerate to one system but not another. Note that this definition of degeneracy need not directly appeal to any cansideration of there dimensional models.

3.2 Unlikely Views?

Even this definition of t degenerate viewpoint it incomplete; basically it lays that generality is a form of stability. Many stable viewpoints ought to be considered degenerate, at least in the sense that that they are less likely to allow a system to instantiate a proper model than other viewpoints.

Consider a pyramid with a square base and arbitrary height (Customarily, it has equilateral triangles for its sides, but we relax mat restriction.) Imaged from many viewpoints from below, its image appears to be a type of rhomboid: a tilting square. None of these viewpoints is degenerate according to he stability definition above, since a slight cnange in viewpoint does not cause a drastic change in the image; it merely tilts the rhomboid In fact, there is a great deal of viewpoint freedom, and many views appear to yield the same semantic result: a partly instantiated pyramid with height information largely missing, what is most disturbing about such views is that they are potentially the most common. For a very flat pyramid, such rhomboids appear from nearly half of all viewing directions.

Yet it seems plausible to suggest that these particular views be considered at least partially degenerate; in contrast to some other views, these images give little information about how to instantiate the pyramid's height (They do place weak upper limits on the height: the peak is constrained to heights that keep it invisible.) Further, if our model base were more complete, we would not be able to distinguish such a view from among similar views of a triangular wedge with square base, or any similarly tapering polyhedron with a square base, despite the stability of our vantage point Thus, we may wish to include in our definition of degeneracy those viewpoints from which "relatively little" three-dimensional information may be obtained, regardless of stability.

Operationally, this aspect of degeneracy can be quantified as the expected number of additional views necessary to disambiguate the object; degenerate$_t$ view is therefore one that is relatively unifororative and will requite mote linages. This number clearly depends on the complexity of the model data base, the intelligence of the system procedures for determining the "best" next view, and the (necessarily) heuristic procedures of the system for inverting the image projection. Again, this is a system performance definition, and an image s degree of degeneracy would change as system parameters change.

It would appear that this definition must be probabilistic. It is not hard to conceive of objects or object assemblies in which multiple viewpoints give identical images, but for which the resolution into a single interpretation takes varying strategic paths depending on the differing image features that can appear in the second view. (Take as an example a cube with a single distinguished face, viewed initially so that only one non-distinguished face is visible.) The likelihood of taking each path can be quantified: the most inclusive measure of an image's degeneracy would be then be its probability-labeled search tree. Various measure* based on the full tree (one of which is, of course, is expected depth) could also serve as s measure of degeneracy. Under this definition, tree breadth has no strong role: a degenerate image can resolve itself into hundreds of images, but as long as each new image was interpretable. the original image is no less degenerate than the pyramid viewed from below.

It is interesting to note a paradoxical consequence of this systems' view of degeneracy. As a system's power increases due to the availability of more sophisticated shape analyzing tools (such as when shape from skewed symmetry is used widh shape from shadins), more types of ambiguity are possible. Each method brings with it a weakness. The implication is that vision systems with multiple sources of knowledge must know when to ignore a source undergoing degeneracy. This meta-knowledge can be explicidy coded, or implicitly handled by means of a flexible enough representation that permits "don't know much" as a valid answer.

4 Specific Imaging Degeneracies

Considering now only images of polyhedral objects, it is possible to eive a catalogue of image degeneracies. Each is based on a specific heuristic for inverting the three-dimension to two-dimension image function. The list is partial, and omits some heuristics that are even more fundamental, such as the generally assumed heuristic rules that lines in the image have been caused by lines in three-space. (In this last case, this would imply that any planar curve imaged from within its plane would often be considered as degenerate, if the system were unable to interpret it as other than a linear object)

4.1 Vertices Imaged in the Plane of Scene Edges

Apparently the major source of "degenerate views" in the blocb world (see Figure 1), so-called coincidental alignments occur when the image of a vertex appears to fall on the image of an edge. They confound the basic imaging assumption that three or more lines coincident in the image are coincident in the scene. If the scene is analyzed using labelling, the labelling will fail. The image is then degenerate because another image is required. In

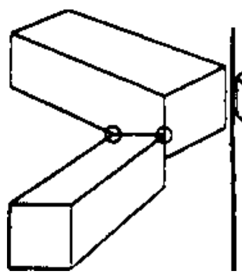

Figure 1 ■ Classical "Degenerate" View

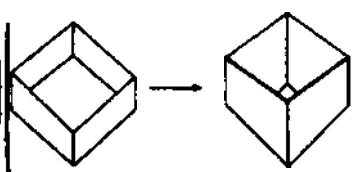

Figure 2

Parallel Scene Lines Imaged in their Own Plane

theory, such coincidental alignments have probability zero, since the camera must lie on a specific plane (or more precisely, in the infinite intersection of two co-planar half-planes).

42 Parallel Scene Lines Imaged in Their Own Plane

This is one of the degeneracies observed with the cube (see Figure 2). It violates the heuristic that colinear in the image implies colinear in the scene, a heuristic often not used. Hence, it is system-sensitive. In *theory,* it also has probability zero, although the camera placement is somewhat more free than in the case of vertex-on-edge.

4.3 Coincident Scene Lines Imaged in Their Own Plane

This is a special case violation of linear in the image implies linear in space. Again, this is system dependent and has probability of zero. (Figure 3 illustrates one example of this degeneracy while viewing a pyramid.)

4.4 Perfect Symmetry

This is an interesting extreme case, and one apparently avoided by professional photographers as it appears to flatten relief (we provide a straightforward example of perfect symmetry in Figure 4). It is apparently based on the heuristic that symmetry in the image implies symmetry in the scene perpendicular to the line of sight. Analogous to what happens when a cube is imaged as two congruent rectangles, it is often degenerate since perfectly symmetric images lack the cues to depth that broken (skewed) symmetries provide. It has a probability of zero of occurring ideally, although the camera now has the freedom to move in at least one entire plane.

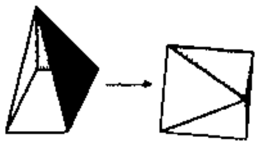

Figure 3

Coincident Scene Lines Imaged in their Own Plane

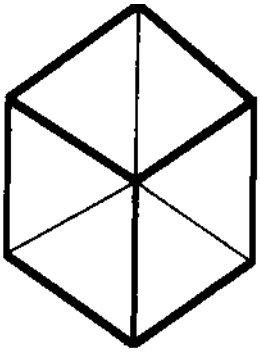

Figure 4

Perfect Symmetry (View of a Cube)

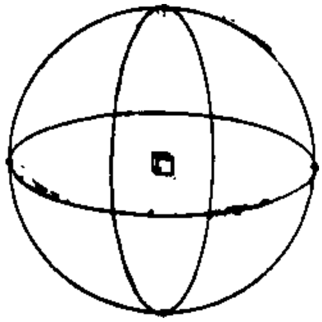

Figure 5

Cube, at Center of the "Viewing Sphere"

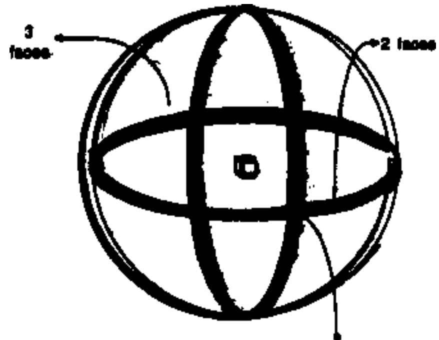

Figure 6

Cube at Center of the Viewing Sphere, Degeneracies Delineated

5 The Effect of Finite Resolution

The various viewing points for our camera may be modeled as points on a viewing sphere at whose center lies the object of interest Therefore, in the ideal case of the cube with infinite resolution under orthography (or, for that matter, under perspective) there are precisely six viewing directions from which we see exactly one face of the cube and no more. Similarly, a family of three mutually orthogonal great circles which intersect at these six points determine the set of directions from which we would see exactly two faces of the cube. Anywhere else on the sphere we see three faces (see Figure 5). A point on the sphere chosen at random will be a viewpoint imaging three faces with probability 1.

Of course, any real system will have only finite resolution. How this resolution is measured, and how repeatable it is, can vary depending on application. For the cube, resolution appears to be the ability to separate two nearly concurrent parallel lines into their separate sources. (Under perspective, the parallel lines would only be nearly parallel.) Assuming this resoivability of parallels is independent of the line segment lengths (admittedly, this is somewhat unrealistic), then the zero probabilities of degeneracy become finite. On the viewing sphere, the great circles have become bands, and the points of single face viewing have become spherical squares (see Figure 6). Their relative areas (and hence the probability of degeneracy) are straightforward to compute in terms of camera acuity. The less accurate the camera, or the farther it is away, or the smaller the object, the larger the likelihood that a viewpoint is degenerate. In the extreme, the bands merge, and no viewpoint sees a "characteristic" view: the images are infinitely degenerate.

The edges of the bands, however, cannot be sharp. Although the bands partition the surface of the viewing sphere, their borders represent those viewing directions at which parallel lines are first" seen as two lines. Given camera inaccuracy and noise, this transition to resoivability cannot be sudden. Depending on the camera and the accuracy of the algorithms processing its data, repeatability may best be representee! by a fuzzy boundary. Thus instead of each point on the sphere having a label, it has a vector of (label, likelihood) pairs. Each degeneracy region, then, fades away in likelihood as it extends farther from its ideal point or great circle. The actual computation of the shape of this probability density depends on the image heuristic used for image function inversion, as well as some measure of camera and software precision. Nevertheless, the relative size of the integral of probability over the partition can give an accurate estimate of the likelihood that a particular view, degenerate or not, will be visible after a random camera placement

5.1 Comments

In a sense these aspect maps are property spheres, where the property is a type of degeneracy. Computing them demands substantial computational time and storage [Besl and Jain 85]; both would benefit from a hierarchic, trixel-like approach. Note that since the sphere it topologicaly equivalent to the extended plane, all such maps can be drawn as planar graphs [Werman, Baugher, and Gualtieri 86]. (See Figure 7, where the "standard"aspect graph has been augmented to show all possible transitions out of degenerate views, as alluded to in [Castore 84]"*).

6 Minimal Disambiguation Distance

Transitions between the various regions of the sphere represent what [Koenderink and van Doorn 79] term a "visual event . Only such a transition is capable of yielding qualitatively new information. Thus these transitions clearly represent a useful change of viewpoint, which would be worth paying for in terms of traveling distance. ,

For the purposes of minimizing the distance traveled in obtaining further images, it will be useful to quantify the minimal distance we must travel on our viewing sphere, to ensure that we will experience a visual event If we reach a characteristic view from which an image has not yet been obtained, then we will maximize the probability mat any degeneracy will be

#**m particular, we have added, without ton of plannarity, the direct transitions between regions where 1 face it visible, and regions where 3 faces am visible. We noticed that in their excellent survey on 3D object recpgnition., [Beil and Jaim 15, p. 19] did not show such transitions in their aspect graph of a cube, nor did their analogues appear in the aspect graph of a tetrahedron, presented by Koenderink and van Doom. For an intersting discussion of the relevance of such transition! to robotic vision, see the remarks of N. Badler in [Castore 84]

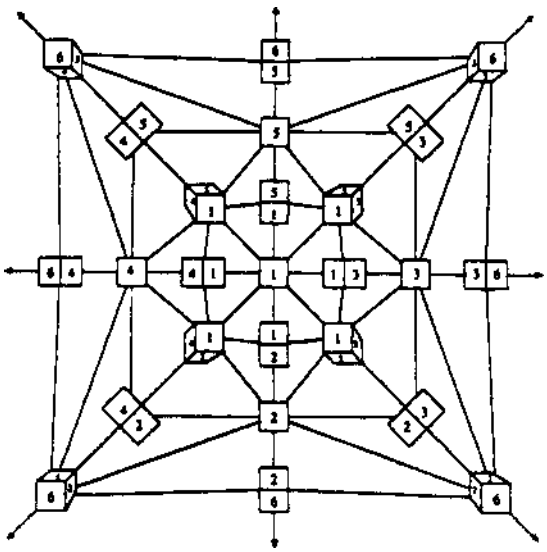

Figure 7

[Planar] Aspect Graph for Cube, including 1-face to 3-face Transitions

disambiguated by the new information. In the case of our viewing sphere of a cube, this would mean ensuring that a 3-faced image is obtained.

Using the assumptions of orthographic projection, we can easily quantify this minimum distance (for the geometric proof, see [Kender and Freudenstein 87]). For a fixed viewing-sphere radius of resolution *n* (i.e. *n* is an absolute number equal to the number of pixels our object will occupy in the image), the distance we must travel along the viewing sphere is equal to the product of the length of the radius and the arcs in of *(1/n)* radians.

In other words, we can quantify the minimal disambiguation *angle* as theta, where

$$\sin\theta = \frac{1}{n}$$

7 Relative Probabilities for the Cube

We might, in a given situation, wish to know the probability of reaching a particular class of viewpoint, given a *random* decision as to "where to go next." For the case of the cube, we can obtain the probabilities of 1, 2, and *3-faced* views, by using spherical geometry to calculate the relative surface areas on the viewing sphere of the 3 regions described above and depicted in figure 5 (square, rectangular, and triangular patches).

An analysis of the cube shows that these probabilities are generally a function of system resolution : Systems capable of higher resolution will generally be less likely to yield "uncharacteristic" views, as one would *expect.*

Our results, the complete derivations of which are presented in [Kender and Freudenstein 87], are as follows:

$$P_{SquarePatch} = \frac{6\,\theta\sin\theta}{\pi}$$

$$P_{RectangularPatch} = \frac{3\sin\theta\,(\pi - 2\theta)}{\pi}$$

$$P_{TriangularPatch} = 1 - 3\sin\theta + \frac{6\,\theta\sin\theta}{\pi}$$

8 Closing Observations and Future Research

Calculating the number of necessary views and the effort to obtain them is a formidable task. In some senses it resembles the design of part feeders [Natarajan 86]: that is. given an unknown position on the viewing sphere, determine what series of camera movements would inevitably lead to a distinguished configuration, namely, the acquisition of all relevant semantic information about an object or object assembly. Even assuming one knows perfectly where one is on the viewing sphere, the determination of even the distance to the nearest visual event is complex, given its probabilistic nature. Circumstances are easy to construct (for example, when the object is too small) where it is actually impossible.

The aspect map can be augmented with other information. It can incorporate probabilities such as the likelihood of a given gravity-induced preferred orientation, or it can be convolved with a placement uncertainty spread function. The spread function can be variable, itself incorporating such information as the robotic work space or other constraints on placement motion. A search for the optimal next view could then also minimize camera placement error and also, by related methods, camera placement costs.

Such algorithms would be particularly valuable if ways exist to formally combine the aspect maps of individual objects to create the aspect map of an object assembly. Thus, from a few primitives and a little knowledge of the robotic placer and its workspace, a single representation could direct active sensing. Whether or not such a representation is ultimately practical, it has nevertheless been helpful in elucidating the meanings of "general viewing position and "degenerate view".

References

[Besl and Jain 85] Besl, F. J., and Jain, R. C. Three-Dimensional Object Recognition. *ACM Computing Surveys,* 1985.

[Castore 84] C as tore, G. M. Solid Modeling, Aspect Graphs, and Robot Vision. In *Solid Modeling by Computers: From Theory to Applications.* Plenum Press, 1984.

[Chakravarty 82] Chakravarty, I. *The Use of Characteristic Views as a Basis for Recognition of Three-Dimensional Objects.* PhD thesis, Rensselaer Polytechnic Institute, 1982.

[Chin and Dyer 86] Chin, R. T., and Dyer, C. R. Model-Based Recognition in Robot Vision. *ACM Computing Surveys,* 18(1):67-108, March, 1986.

[Kanade80] Kanade,T. A Theory of Origami World. *Artificial Intelligence,* 13(3):279-311,May, 1980.

[Kender and Freudenstein 87] Kender, J. R. and Freudenstein, D. G. What is a Degenerate View? In *Proceedings of the DARPA Image Understanding Workshop,* pages 589-598. Los Angeles, CA, February, 1987.

[Koenderink and van Doom 79] Koenderink, J. J., and van Doom, A. J. Internal Representation of Shape with Respect to Vision. *Biological Cybernetics,* 1979.

[Lavin 74] Lavin, M. A. *An Application of Line-Labeling and other Scene-Analysis Techniques to the Problem of Hidden-Line Removal.* Technical Report, MIT Artificial Intelligence Lab, 1974.

[Natarajan 86] Natarajan, B.K. *On Moving and Orienting Objects.* PhD thesis, Cornell University, 1986.

[Sabbah82] Sabbah, D. *A Connectionist Approach to Visual Recognition.* PhD thesis, Rochester University, 1982.

[Thorpe and Shafer 83] Thorpe, C, and Shafer, S. *Topological Correspondence in Line Drawings of Multiple Views of Objects.* Technical Report, Department of Computer Science, Carnegie-Mellon University, 1983.

[Werman, Baugher, and Gualtieri 86] Werman, M., Baugher, S., and Gualtieri, J. A. *The Convex Visual Potential.* Technical Report, University of Maryland, Center for Automation Research, 1986.