

# THE INTERPRETATION OF THREE-DIMENSIONAL STRUCTURE FROM IMAGE CURVES

David G. Lowe and Thomas O. Binford

Artificial Intelligence Laboratory, Computer Science Department  
Stanford University, Stanford, California 94305

## Abstract

*General constraints on the interpretation of image curves are described and implemented. We illustrate the use of these constraints to interpret three dimensional structure from an image up to the volumetric level. These constraints do not use any knowledge regarding the specific objects in the image, but rely on general assumptions regarding illumination, occlusion, object geometry, and the imaging process. They are based on coincidence assumptions that various coincidental features or alignments in an image are unlikely to arise without reason. The strength of these coincidence assumptions depends on the accuracy of the low-level description of an image. Since any one feature can be the result of pure coincidence or errors in the detection process, it is necessary to use a reasoning system which can use these hypotheses to derive consistent global interpretations, while maintaining the ability to remove the implications of hypotheses which are disproved in the face of further evidence. We demonstrate the successful interpretation of some hand derived image curves up to the volumetric level, including the construction of a three-space model.*

## Introduction

An image can loosely be thought of as a two-dimensional projection of the light reflected from three dimensional physical objects. Since most physical objects in our world have sharp physical boundaries (as a condition for maintaining their physical integrity), it is common for the visible edges of objects to appear as sharp intensity discontinuities in an image. Because these discontinuities are sharp, it is possible to locate the position and orientation of these curves accurately in the image. For this and other reasons, we believe that intensity discontinuities (also known as image curves or lines) are a particularly valuable indicator in an image of an object's shape and geometric structure. Current work in neurophysiology [4] indicates that the mammalian visual system also places considerable emphasis on the detection of intensity discontinuities. This paper is concerned with the use of these two-dimensional curves to derive the three-dimensional geometric structure of the objects in an image.

Previous work on the interpretation of image curves, such as that by Huffman [5], Clowes [3], Waltz [8], and Kanade [6], has usually concentrated on the constraints imposed on boundary junctions by certain classes of geometric objects. We believe that there are more general constraints on the interpretation of image curves, and that the use of these constraints is of great importance for the interpretation of real data and general classes of images. Most previous attempts at boundary interpretation have used connectivity as the main source of information, whereas we have placed at least as much emphasis on location, shape, size and

orientation. Unlike most previous work, our implementation of these constraints is designed so that no single constraint is taken to be absolute, and the system can therefore tolerate some errors and incompleteness in its initial data.

The constraints we use are based on coincidence (genericity) assumptions, which require no prior knowledge about the specific objects in the image. Many types of alignments in an image are unlikely to arise by coincidence, given that we can reasonably assume a general camera position and general position of the sources of illumination. For example, when three or more curves converge to the same vertex in the image, it is either a coincidence in which the camera is restrictively placed to produce the alignment, or the three curves meet at a common vertex in three-space. It is reasonable to base our interpretation on the non-coincidental case, until there is other evidence that a coincidence has in fact occurred.

By applying constraints such as these, we can categorise image curves into three distinct classes: those caused by discontinuities in the geometry of an object (edges), in the reflectance of an object (markings), and in the illumination (shadows). This categorization in turn leads to other constraints on the geometry and surface structure of the objects, their occlusion by other surfaces, and to a layered ordering of the surfaces in terms of their distance from the camera. Where geometric edges cast shadows, it is usually possible to match the edge to the shadow and make precise measurements of the relative separation of surfaces.

The constraints described in this paper are meant to interface with both higher and lower levels of visual interpretation. This work is being done in the context of the ACRONYM vision system [2], and is expected to interface closely with the higher levels of ACRONYM, which apply constraints derived from generic object models. The strength of the coincidence assumptions described here depend strongly on the accuracy of the low-level curve descriptions, and in turn, many of the constraints can lead to better descriptions of the shape and intersections of the curves. Although the only examples shown in this paper are based on hand-derived curve data, work is currently under way to interface these constraints with new curve detection techniques being developed as part of ACRONYM [1, 7].

## Constraints on the interpretation of image curves

This section describes a series of general constraints on the three space interpretation of image lines. In each case we attempt to describe the specific coincidences which would lead to alternate interpretations. Many of these constraints are also described in [1].

In a certain trivial sense, any of the image features described in the following constraints could be the result entirely of reflectance discontinuities rather than geometric or illumination

discontinuities, since the image could be a picture of a picture. However, we prefer interpretations which are consistent with solid geometry, and assume coincidentally similar surface markings only in the face of further evidence.

1) Alignment of image curves. When two straight lines are aligned in an image (even though they may be separated by a gap of a considerable distance), they must also be aligned in space, or else the lines must be parallel and the observer must be coincidentally in the plane of the two lines. This constraint allows us, for example, to hypothesize continuity of line segments on both sides of an occluding object. This constraint can be extended to deal with the alignment, of circular arcs, smooth splines, symmetries, and any shapes which are predicted from knowledge of the imaged object. It can also be used to bridge gaps in curves due to errors in the curve detection process or insufficient contrast across some part of a boundary. By combining a model of the errors in the curve detection process with the separation and support for the aligned curves, it is possible to quantify the strength of the constraint in each individual case. This is an area of current research while we are integrating these constraints with the curve detection and description process.

2) Tangent discontinuities in curves. Tangent breaks in an image curve (also known as tangent discontinuities or corners) are important for the same reason that intensity discontinuities are important in an image: they are features which can be sharply localized, and can therefore lead to more restrictive constraints than more diffuse features. Two-dimensional breaks in an image curve are closely related to breaks in three-dimensional space curves: in fact, a smooth space curve will never have a break in its image. Likewise, a break in a space curve will result in a break in its image, unless there is a coincidence in which the observer is coplanar with the two tangents. The situation is the same for a cast shadow: there will be a break in the shadow cast by a geometric edge which contains a break, unless the source of illumination is coincidentally coplanar with the two tangents of the break.

3) Straight lines. An image line which is straight must be the image of a straight space curve, unless the curve is planar and the observer is coincidentally aligned with the plane of curvature. Likewise, a straight shadow curve in the image must have been cast by a straight geometric boundary onto a planar surface, unless the source of illumination is coincidentally in the plane of curvature of the geometric boundary or the observer is in the plane of curvature of the shadow curve.

4) Termination at a continuous curve. When an image curve terminates at a continuous curve (a T junction), the terminating curve cannot be closer to the observer than the continuous curve; otherwise, it would be a coincidence that the termination happened to occur on the other line. The T junction could be the result of three different occurrences: occlusion of the terminating edge by a geometric boundary, the termination of a surface marking at the visible edge of a geometric object, or a specific set of surface markings. Therefore, if we know that the terminating curve is a geometric boundary, then we can infer that the continuous curve is also a geometric boundary and we know its direction of occlusion. If we know that the continuous curve is a geometric boundary occluding on the side of the terminating curve, then we can infer that the terminating curve must be a surface marking. If we know that the terminating curve is a shadow, then we can infer that the continuous curve is a non-concave geometric boundary (if it was concave we would see a continuation of the shadow from the same vertex).

5) Crossing of continuous curves. When two continuous curves cross one another (an X junction), this is an indicator of either an illumination discontinuity, transparency, or an unusual combination of surface markings, since the curve closes! to the observer does not occlude either side of the other (we are assuming that "wires" will be distinguished on the basis of other local evidence, or else they can be handled as a separate case). If one of the curves is an illumination discontinuity (shadow boundary) this implies that the other curve lies on the same surface and is not a geometric boundary, since otherwise a break would occur (as will be discussed further below).

6) Contrast across shadow edges. The contrast across a shadow edge is equal to the ratio of direct to indirect light, which remains fairly constant across an image for a distant light source. However, the situation is complicated by the presence of reflections from nearby objects, which can considerably increase the amount of indirect illumination. A stronger constraint, is that the contrast ratio along the length of a shadow will change only smoothly and independently from the surface on which it falls, so that as the shadow curve crosses different, reflectance boundaries there will be the same contrast ratio on both sides of the boundary. This is a valuable constraint for identifying shadow lines. In the case of the X junction mentioned above, the illumination discontinuity will have the same contrast ratio on both sides of the junction, and can therefore be distinguished from the other curve at the X junction, barring coincidence. In color imagery, the ratios of illumination at each wavelength across a shadow boundary should be fairly constant and vary only smoothly along the length of the shadow, and could therefore provide an even stronger constraint (eg. the dark side of a shadow boundary will be bluer than the sunlit side on a clear day).

7) Shadow breaks caused by geometric breaks. As mentioned in (2), barring coincidence, breaks (tangent discontinuities) in geometric edges are observable as breaks in image lines, and geometric edges with breaks cast shadows with breaks. Therefore, given a known direction of illumination, the geometric break causing any particular shadow break is constrained to be in a precise direction. This constraint is so strong that the mere existence of a break in the given direction is support for the hypothesis of a shadow. If the direction of illumination is unknown, then sets of matching breaks between hypothesized shadows and other image lines, which converge to a single point (see next paragraph), provide evidence for the hypothesis of the direction of illumination.

In perspective imagery there is an *illumination convergence point* in the image through which the images of all illumination rays from a point source pass (this is true even for nearby point sources). The location of this point can be determined from any set of two or more matches between shadow and geometric breaks. If the point source is in front of the camera lens plane, then the convergence point is of course the location of the image of the point source. If the light source is behind the camera lens plane, then the illumination convergence point is located at the point of projection of the light source onto the film plane through the projective center of the camera, and the illumination streams towards this point rather than away from it. If the point source is exactly in the lens plane of the camera, then the perspective effect compensates for divergence from the light source to make the illumination convergence point infinitely far off, so all images of illumination rays are exactly parallel (these insights are due to Sid Liebes).

It should be noted that breaks in shadow curves can also be

caused by the intersection of shadows cast by different objects, so it cannot be assumed that there is a geometric break corresponding to every shadow break. The constraints given here could be extended to cover discontinuities in curvature and other shape properties as well as discontinuities in tangents, although these are more difficult to implement computationally.

8) Casting of shadow curves by geometric edges. Given one or more matches between shadow breaks and geometric breaks it is easy to determine which geometric edges are casting which shadow edges. An important constraint is that each point on the shadow edge must correspond to some point on the geometric edge in the direction toward or away from the illumination vanishing point, although all parts of the geometric edge may not be observable (say, if it is occluded by another object). Once a match has been hypothesized between a shadow edge and a geometric edge, many important inferences follow, the casting curve is known to be a geometric edge or limb with the direction of occlusion depending on the direction of contrast across the shadow edge. If the geometric edge is straight then any curvature in the shadow curve is due to curvature in the surface on which it is cast. Likewise, if the shadow surface is known to be planar, the curvature of the shadow can be used to calculate the curvature of the casting edge. The image separation of the casting curve from the shadow can be used to calculate their relative range from the camera. Shadows essentially provide a second projection of the object geometry, in addition to the original image, and can therefore be used in much the same way as stereo information.

9) Junctions of two or more discontinuous curves. When two or more curves terminate at the same junction, it would be a coincidence if the observer were aligned so that separated vertices in space landed at the same point in the image. Therefore, for L, Y, K, or higher order junctions, it is reasonable to hypothesize that coincidence in the image implies coincidence in space. In other words, knowing constraints on the 3 space location of the endpoint of any of the terminating curves gives us the same constraints on the 3 space locations of the other endpoints at the vertex. When a shadow curve terminates at a Y or higher order junction, we can assume that one of the other terminating curves is a geometric edge casting this shadow onto a surface passing through the junction (otherwise it would be a coincidence that the shadow happened to pass through the junction).

10) Propagation of direction of occlusion. When the direction of occlusion is known for a geometric boundary (ie., we know which of the surfaces on either side the edge belongs to), then any unambiguous continuations of the geometric boundary will have the same direction of occlusion.

11) Surface continuity. We assume smoothness and continuity of surfaces when there is no intensity discontinuity in the image. If an intensity discontinuity on a surface is due to a geometric discontinuity of the surface, then we would expect to see a discontinuity in the geometric boundary of the surface where it intersected this intensity discontinuity (unless the observer is in the plane of the two tangents at the boundary). Therefore, a curve which intersects a smooth boundary of the surface on which it lies can be assumed to be a surface marking rather than a tangent break in the surface.

Knowing the direction of occlusion (from 4,8,10, etc.) allows us to form surface descriptions by looking at the regions bounded by geometric boundaries. In many cases it is possible to form surface descriptions by merely following the continuations of geometric boundaries around the extent of the surface, until they

are possibly occluded by another surface. Another technique is to work away from the occluding side of a geometric boundary, ignoring illumination and reflectance discontinuities, until either an opposing geometric boundary is found or an occluding geometric object is found. Note that, this is different than the usual region segmentation of images, since we are doing it in the geometric domain. These can be difficult constraints to implement computationally, since they deal with properties of entire surfaces rather than just a curve. Our current implementation handles only some of the more common cases of surface description.

12) Prediction of shadow locations. If we have formed hypotheses of the geometry of an object, the surrounding surfaces, and the direction of illumination, then we can predict the locations of illumination discontinuities in the image, and thereby produce new hypotheses for image lines at these locations, as well as confirm or contradict our original hypotheses. Prediction of this sort can also be used to check consistency of surface occlusion.

13) Parallelism in object space. In perspective imagery, straight lines which are parallel in object space all converge towards a common vanishing point in the image. Since scenes often contain many lines parallel to a preferred direction (such as vertical with respect to the ground), once a vanishing point has been identified for some lines, other lines which are aligned with it can be assumed to be parallel in object space, barring coincidence. Curves which are parallel and close together in object space (relative to the image size), are almost parallel in the image for most camera orientations (this can be precisely quantified).

Since we are attempting to derive three-dimensional structure from image clues, the constraints above all deal with image features which are *quasi invariant* (ie., generically stable, with at most a planar locus of instability) with respect to viewpoint. For example, a curve break in three space produces an image curve with a break over a wide range of viewing conditions, even though the angle of the break is not invariant. The higher levels of ACRO-NYM produce predictions of quasi-invariants from specific object models, and these should interface well with the more general quasi-invariants described above.

The list of constraints given above can be extended in many directions, while still maintaining the same level of generality. We are particularly interested in constraints for inferring the cross sections and volume descriptions of geometric objects in an image.

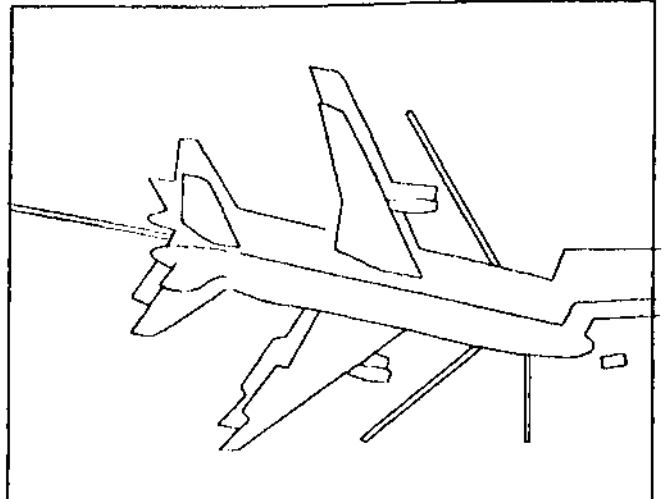
#### Reasoning on the basis of coincidence assumptions

The use of incomplete constraints forces us to use a non-monotonic reasoning system, in which further evidence can disprove a previously held hypothesis. We have chosen a method of deduction which forms an explicit symbolic representation of each hypothesis, and maintains a record of its support and implications so that it can be reevaluated and undone if new contextual evidence indicates that the original hypothesis was false. Each hypothesis has pointers to all the hypotheses and types of evidence which have given support to it, as well as pointers to the hypotheses which it supports.

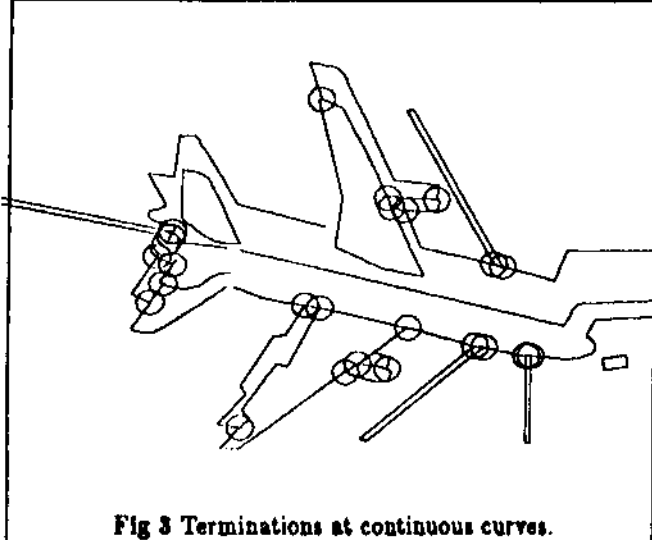
When conflicting interpretations are suggested for some part of the image, the evidence for each interpretation is evaluated to see if one interpretation is strong enough to exclude the other. If there is insufficient information to make this decision, then both hypotheses are pursued until there is. When a previously held hypothesis is thought to be false, its implications are undone and any further results propagated. This non-committal style of deduction works well because the constraints are strong and the



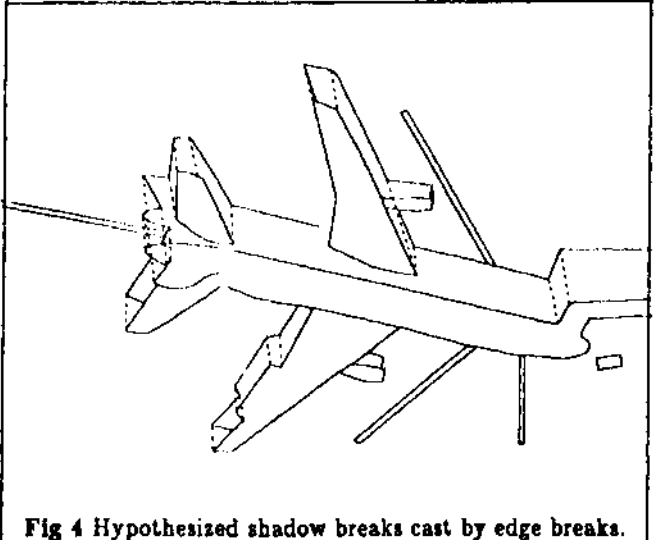
**Fig 1** Original digitized image.



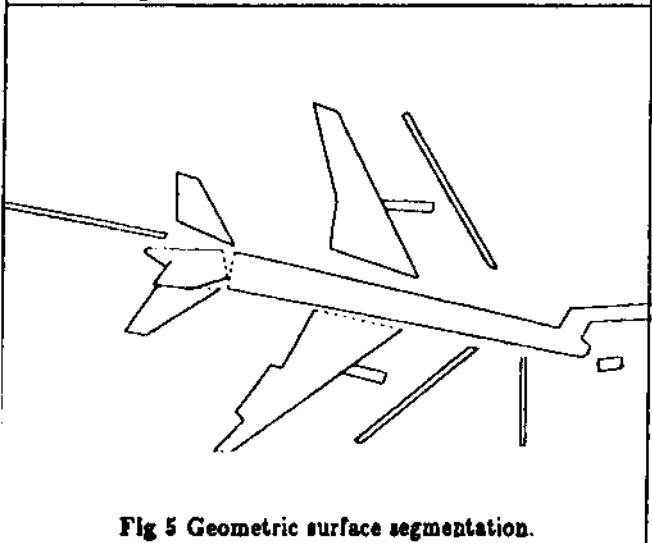
**Fig 2** Hand derived curve data used as input.



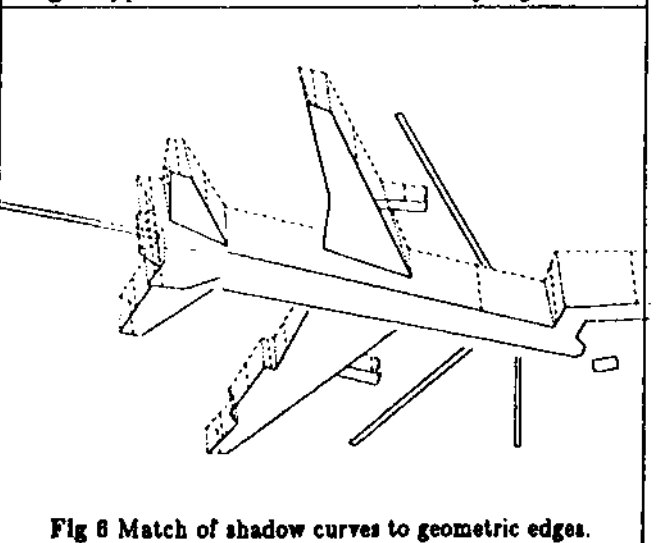
**Fig 3** Terminations at continuous curves.



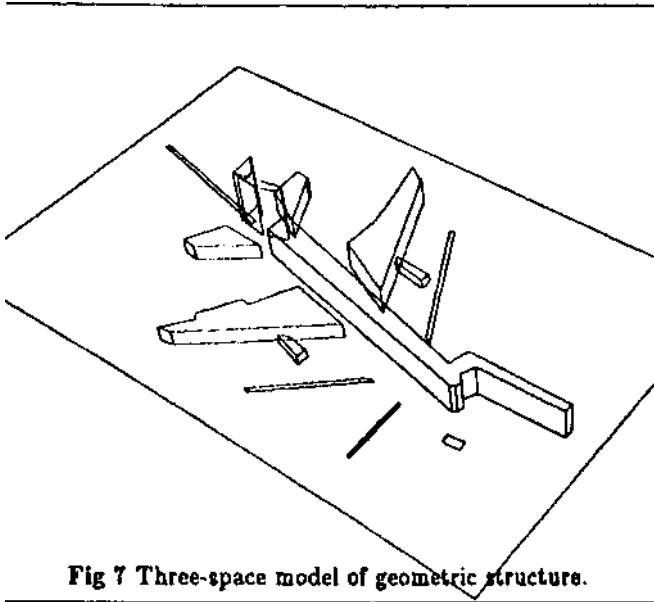
**Fig 4** Hypothesized shadow breaks cast by edge breaks.



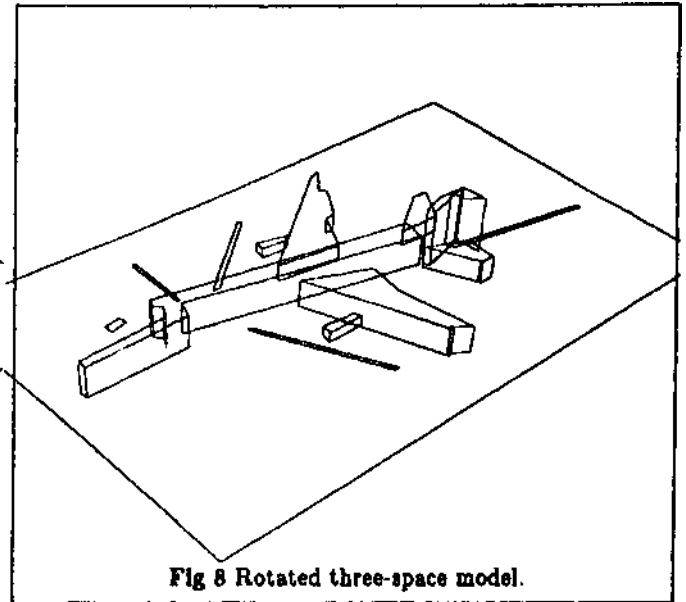
**Fig 5** Geometric surface segmentation.



**Fig 6** Match of shadow curves to geometric edges.



**Fig 7 Three-space model of geometric structure.**



**Fig 8 Rotated three-space model.**

search space is shallow. The major difficulty is that this requires that special code be written to evaluate and resolve each type of conflict that could occur, but we are looking at other possible resolution schemes. The explicit maintenance of constraints is much better than the use of backtracking for testing hypotheses, since it is not limited to the last-in-first-out order of testing typically enforced by a backtracking stack, and any results apply to the entire problem space.

The coincidence assumptions could sometimes be used to measure the likelihood that a given coincidence will occur, and therefore be used to compute a probabilistic degree of confidence in a hypothesis. However, these probabilistic measures can be extremely context-dependent and may vary widely from image to image in ways that can not be known before the image is interpreted (i.e., the probabilities cannot reasonably be assumed to be independent). The combination of probability values may be of importance in some cases for speeding up the analysis, by telling us what to examine first, but we have made no attempt to use them in our current system. When some consistent overall interpretation has been found for an image, there is usually much redundant information available in support of the interpretation and there is no need to choose between alternatives on the basis of explicit probabilities.

#### An Implementation

We have written a preliminary version of a computer program which implements the reasoning system and many of the constraints described above. This program has been tested on simulated image curve data derived by hand from a real image, with the encouraging results described below. We are in the process of implementing more constraints, and hope to soon test the program on data derived automatically from images by new curve detection techniques being developed here at Stanford. The program has been implemented in MACLISP on a DEC KL-10, using the record package and ACRONYM environment created by Rod Brooks [2]. All of the constraints we currently use are computationally inexpensive, and the hypothesis formation for the given example took less than 2 seconds of computer time.

The initial curve data is translated into a set of "curve hypotheses" in order to initialize the hypothesis generation process. Each curve is represented as a series of points with the tangent of the curve given at each point, and cubic splines are assumed as the method of interpolating for intermediate points. All curves are indexed in a grid array under all grid squares through which they pass, making it economical to search image neighborhoods for curve features. During input, each curve termination is linked to other curves which terminate or pass through the same spot, as well as to curves which are aligned with its continuation.

A weakness of the current system is its control structure. Currently, most of the hypothesis generation proceeds in a fairly fixed order, with the evidence which is expected to be the most reliable being examined first. However, we are in the process of creating a more flexible control structure using a generalised agenda for ordering constraint propagation.

In spite of the incomplete state of the current system, it is able to carry out detailed interpretations from simulated curve data as shown in Figures 1 through 8. Figure 1 shows the original picture taken over San Francisco airport from which the curve data in Figure 2 was derived by hand. This data also specifies the approximate contrast across edges. In deriving this curve data we referred to data derived automatically by curve detection programs, and we believe data of at least this quality will soon be available. We have also input the location of the sun, although we hope to soon be able to derive the illumination direction(s) automatically from the image. Figure 3 shows the hypotheses regarding curve terminations at a continuous curve (constraint

<sup>4)</sup> Figure 4 shows a dotted line at the match between each hypothesized shadow discontinuity and its corresponding edge discontinuity in the sun direction. This makes a small amount of use of the input data regarding the contrast across curves (constraint 6), but the major support for a shadow hypothesis is the existence of one of these shadow matches (constraint 7). We expect to be able to use tighter tolerances when using automatically generated data than were used with this simulated

data, with correspondingly better results. Figure 6 shows curves which have been hypothesized to be shadows as dotted lines, and also the hypothesized connections between shadow lines and the geometric edges which cast them (constraint 8). These hypotheses were formed from the evidence of Figure 4 in combination with other projective constraints mentioned in the previous section and hypotheses regarding which edges constituted geometric boundaries. Figure 5 shows hypotheses which form surface boundaries from other hypotheses regarding geometric edges, the direction of occlusion, and unambiguous boundary continuations.

From the results of this interpretive process, it is possible to construct three-dimensional models of the scene. The shadow to curve matches provide information on the relative distance of each geometric edge and shadow surface from the camera. Even if the sun position is not known, all measurements are correct relative to some constant factor and can be made absolute by knowing the correct measurement for any one part of the image. Figures 7 and 8 show different views of such a model generated automatically from the hypotheses described above. Some of the minor errors in this model are actually artifacts of the display system rather than being present in the actual hypotheses. The position of the camera with respect to the ground was given to the modeling system (so that verticals would appear vertical), but this could be deduced from the image if the orientation of any image object with respect to the ground was known. With our digitized data it was impossible to see into shadows on the ground because of digitising limitations, even though this information was available in the original negative. We believe that interpretation would have been easier with more complete data. This example by no means exhausts the information inherent in the given curve data, even without considering the use of specific knowledge regarding airplanes and airports.

#### Conclusions

The constraints on curve interpretations do not need to rely on restrictive assumptions about the class of objects in the image. One objective of this research has been to show that image intensity discontinuities carry far more information than has often been recognized in the past. The ability to precisely locate features is of great importance for providing strong constraints on interpretation. Not only is it valuable to localize intensity discontinuities as accurately as possible, but it is important to localise curve terminations, intersections, and tangent or curvature discontinuities accurately. Much work remains to be done on error analysis with each type of constraint in order to make maximum use of the available information.

More weakly localisable information, such as shading, intensity, and color, while being very important, is usually dependent on other assumptions regarding surface continuity. We believe that it is valuable to first examine the intensity discontinuities and carry out interpretation similar to that described here as a step towards using these other sources of information.

Shadows are not troublesome features to be removed from an image, but are in fact one of the most reliable sources of low-level information. While some of our constraints on shadows depend on properties of the sun and assume a known location of the light source, we believe that human perception in general makes use of the more qualitative aspects of our constraints on the interpretation of illumination. In particular, the behavior of illumination boundaries as they cross surface markings and geometric boundaries and terminate at geometric vertices provides a great deal of information even under unknown conditions of illumination.

The identification of illumination boundaries can be carried out at a low level by noting the width of boundary transitions, constraints on contrast changes along a shadow edge, and the crossing of continuous curves where a shadow falls across a reflectance discontinuity.

The reasoning system used for this interpretation task can be summarized as follows: *we reason forward on the assumption that no coincidences have occurred, but always maintain enough information so that hypotheses can be reevaluated and changes propagated in the face of new contextual evidence*. This gives us the ability to use constraints which will sometimes be false, and therefore allows us to use many sources of information which would be unacceptable in a strictly deductive inference or constraint system. This method relies on the large amount of redundant information usually available for vision tasks. The further development of tools for reasoning with suggestive evidence is an important goal for many areas of artificial intelligence research.

#### Acknowledgements

Our task would have been much more difficult without the extensive systems and graphics features created by Rod Brooks during his work on ACRONYM. Sid Liebes has helped UB work out the geometry of perspective imagery and shadow formation. Peter Blicher, Dave Manmont, and Barry Soroka made helpful comments on a first draft, of this paper.

This research has been supported by the National Science Foundation under grant DAR78-15914, and by the Defense Advanced Research Projects Agency under contract MDA903-80-C-0102. The first author also receives support from a National Science and Engineering Research Council of Canada Postgraduate Scholarship.

#### References

- [1] Binford, Thomas O. "Inferring Surfaces from Images" *To be published in AI Journal*, 1981.
- [2] Brooks, Rodney A., Russell Greiner and Thomas O. Binford, "The ACRONYM Model-Based Vision System," *Proc. of IJCAJ 79*, Tokyo, Aug. 1979, 105-113.
- [3] Clowes, M.B. "On seeing things," *Artificial Intelligence*, 2 (1971), 79-116.
- [4] Hubel, David H. and Torsten Wiesel "Brain Mechanisms of Vision" *Scientific American*, September 1979, 150-162.
- [5] Huffman, D.A. "Impossible ObjectB as Nonsense Sentences," *Machine Intelligence*, 6 (1971), 295-323.
- [6] Kanade, Takeo "A Theory of Origami World," *Artificial Intelligence*, 13 (1980), 279-311.
- [7] MacVicar-Whelan, P.J. and T.O. Binford, "Line Finding with Subpixel Precision" *This proceedings*
- [8] Waltz, D. "Understanding Line Drawings of Scenes with Shadows," *The Psychology of Computer Vision*, P.H. Winston, (ed); McGraw-Hill, 1975.