Proceedings of the *IEEE International Symposium on Intelligent Robotics*, Jan 7-9, 1993, Bangalore, India.

Recovering 3-D Shapes from 2-D Line-Drawings

Pawan Sinha
Department of Computer Science and the Artificial Intelligence Laboratory
Room NE43-753, 545 Technology Square

Edward Adelson
Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

A central problem in vision is the recovery of the 3D structure of a scene from a single 2D projection. The interpretation of planar line-drawings represents one of the most challenging instances of this highly underconstrained problem. Quite intriguingly, the human visual system can readily accomplish this task. This paper reports on an attempt to understand and computationally mimic this important perceptual ability. It has long been suggested that humans use principles of simplicity in achieving this percept. Barrow and Tenenbaum, and more recently, Marill, have proposed that standard deviation of the included angles should be used as a measure of complexity; minimizing this metric leads to perceptually correct interpretations for many line-drawings. However, we have found that the use of this measure alone results in unexpected and bizarre interpretations for certain figures. We have devised a new approach that utilizes three types of measures: angle variance, planarity of faces and overall compactness. When these measures are appropriately combined, the model interprets a wide variety of line-drawings of polyhedral objects in a manner consistent with human perception. Our model is also robust in the sense that it works without the need for parameter tweaking to handle different cases.

INTRODUCTION:

The central task of any vision system, whether biological or artificial, is to infer the true nature of a scene given only its two dimensional projection. By the 'true nature' of a scene we imply such properties as relative depth ordering of objects, surface reflectance distribution and the illumination patterns over the scene. The recovery of all these properties is intimately dependent on the recovery of the three dimensional structure of the scene.

The problem of recovering the 3D structure of an object given a single 2D projection of the same is usually highly underconstrained. In some situations, additional cues like shading and texture gradients can be used to aid the recovery process [Horn, 1975; Witkin, 1981; Yuille, 1987]. The interpretation of line-drawings, which are highly impoverished representations of objects, lacking in all cues except for the x, y coordinates of points along the contours (see figure 1(a)) presents a more difficult challenge.

The human visual system seems to effortlessly reconstruct an object in three dimensions given only its 2D line-drawing (see figure 1(b)). This ability is particularly intriguing when one considers the fact that any planar line-drawing is geometrically consistent with infinitely many 3D structures, as shown schematically in figure 1(c). What distinguishes the 'correct' 3D structure from the rest? How might a vision system search for and find the 'correct' structure in the infinite search space of all possible 3D structures consistent with the given line-drawing? These are the two questions that we shall be primarily concerned with in this paper. To keep the scope of our study manageable, we shall consider only line drawings of arbitrary polyhedral objects.

BACKGROUND:

The issue of how a vision system might recover the 3-D structure corresponding to a 2-D line drawing has been extensively studied. Among the most notable early contributors to this field are Roberts, Huffman, Clowes, Waltz, and Barrow and Tenenbaum [Roberts, 1965; Huffman, 1971; Clowes, 1971; Waltz, 1972; Barrow & Tenenbaum, 1981]. The initial emphasis was on recovering the qualitative shape from line drawings. This involved labelling the observed lines as being either convex, concave or occluding. Mackworth [1973] showed that for polyhedral objects, a labelled line drawing is equivalent to a surface orientation graph (SOG) in gradient space (a SOG has vertices that represent the orientations of surfaces and edges that represent their mutual connectivities). However, the precise scale and location of the SOG in gradient space can not be recovered from merely a labelled line drawing. The next step is to recover the quantitative or precise shape of the object represented by the given line-drawing. Kanade [1981] used ideas of skew symmetry to recover shape in some simple cases. Barrow and Tenenbaum [1981] proposed the notion of maximizing symmetry as a metric for recovering shape.

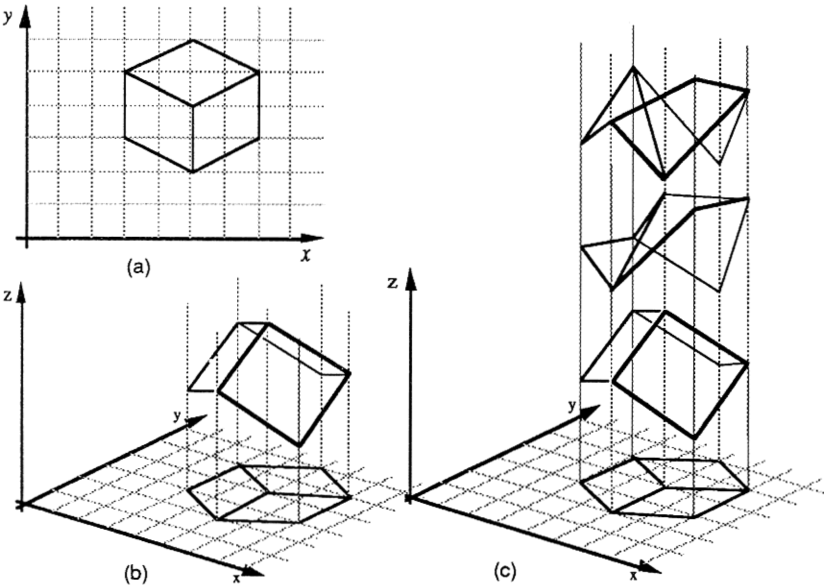

Figure 1. (a) A line drawing provides information only about the x, y coordinates of points lying along the object contours. (b) The human visual system is usually able to reconstruct an object in three dimensions given only a single 2D projection (c) Any planar line-drawing is geometrically consistent with infinitely many 3D structures.

Quantitative shape recovery can be thought of as a transformation of the SOG involving translation and scaling. The required transformation may be determined by the use of some heuristics like those of symmetry mentioned above. Alternatively, the problem may be thought of as one that requires the computation of the unknown z values associated with each vertex visible in the line drawing. This view immediately prompts a 'beads-on-wires' paradigm for shape recovery as shown in figure 2. The idea here is the following: the 'wires' on which the 'beads' rest are oriented parallel to the z-axis (which here is the same as the line of sight). Moving the beads along the wires corresponds to varying the depth values associated with the vertices of the input line drawing. The objective is to choose that configuration of the beads from among the infinitely many possible which minimizes some predefined cost function. Accordingly, the two critical components of a beads on wires implementation are: (1) the objective function, and (2) the technique used to search the space of possible bead configurations. We next review some of the related work that has addressed these issues in the past.

Psychophysical researchers have long suggested that notions of maximizing 'simplicity' underlie many of our visual functions. According to the Gestalt school of thought, there exist global 'principles of organization' that explain the nature of psychological processes. One such principle is the law of Pragnanz, which postulates that 'psychological organization will always be as 'good' as the prevailing conditions allow' [Koffka, 1935]. What exactly constitutes 'simplicity' or 'goodness' is described in a relatively informal and vague manner. It was only very recently that researchers started experimenting with precisely computable measures that could be used as metrics for figural simplicity. Attneave and Frost [1969] interpreted Pragnanz as 'minimum complexity', and suggested that the perception of line drawings of parallelipipeds could be understood in terms of a system that minimizes diversity in angles, lengths and slopes. On similar lines, Barrow and Tenenbaum [1981] suggested some global measures of regularity such as the sum of the squares of angles of faces; the sum of the squares of (2π - Σ(angles at a vertex)) (which tends to equalize an analog of Gaussian curvature at the vertices); the sum of squares of cosines of face angles (which tends to produce right angles). Brady and Yuille [1983] proposed the use of 'extremum principles' in establishing 3-D shape. One such extremum principle is to maximize the ratio of the area to the square of the perimeter. This principle interprets ellipses, parallelograms and triangles as slanted circles, squares and equilateral triangles, respectively.

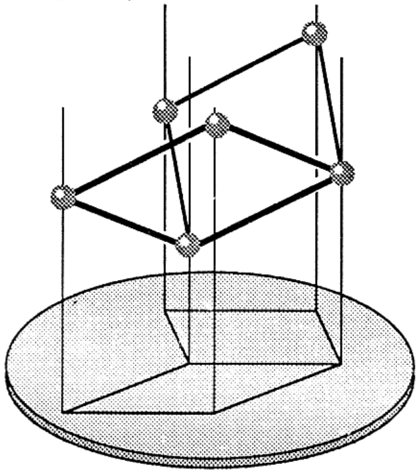

Figure 2. The 'beads-on-wires' paradigm for shape recovery. The basic idea here is to find a configuration for the beads that best satisfies some chosen criterion.

Most of the ideas described above emphasize the need to maximize regularity. A concise computational metric of the psychological notion of regularity is the variance of the included angles in a figure. Minimizing this quantity is akin to making the figure as regular as possible. In a recent paper, Marill [1991] demonstrated that the principle of minimizing the standard deviation of angles (MSDA, for short) can be used to correctly interpret (in a manner consistent with human perception) an impressive array of examples. A simple gradient descent algorithm was used to find the solution that minimized the objective function.

However, further experimentation with the MSDA principle reveals that not withstanding its impressive performance on some examples, it is not adequate for other seemingly simple figures. Figure 3 shows two such instances. The interpretations that the MSDA comes up with are strange objects with warped surfaces. What is interesting to note is the fact that even for incorrect interpretations, the bead configuration usually does represent a global minimum in the standard deviation space. For example, in the hexagonal figure (figure 3(b)), *all* the angles are very close to 90 degrees and the wireframe is maximally symmetric according to the MSDA principle. However, the interpretation does not quite match the natural one - that of a truncated hexagonal pyramid. It is evident that the human visual system does not perform a simple minimization of the SDA. A key difference between these interpretations and the perceptually correct ones is the planarity of surfaces; while all facets in the perceptually correct interpretations are planar, most of the surfaces in the structures shown in figures 3(a) and 3(b) are warped.

Several more examples such as these suggest that planarity is a strong constraint on the interpretations permitted by the human visual system. The human visual system apparently prefers interpretations that are as symmetric as possible under the constraint that all the facets be maximally planar. Another constraint that seems to guide the human recovery of 3-D shape is that of 'compactness'. The idea is illustrated in figure 4. As a direct outcome of the generic viewpoint assumption (the viewing position is not assumed to be distinguished in any fashion), the amount of foreshortening on most of the lines in the 2-D line drawings is not assumed to be excessive. This constraint leads to choosing interpretation 4(a) over 4(b) even though the latter has marginally lesser standard deviation than the former.

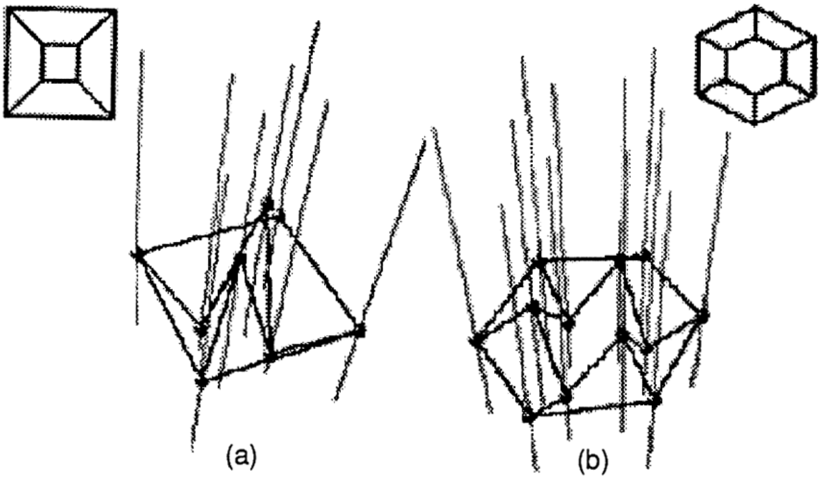

Figure 3. The perceptually incorrect shapes recovered from the input line-drawings (upper left and right) by Marill's algorithm.

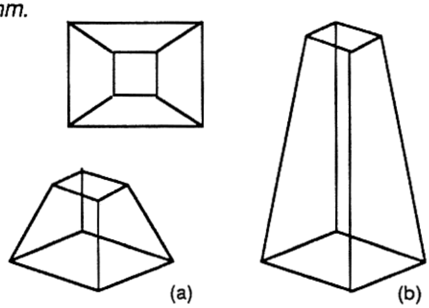

Figure 4. The 'compactness' constraint leads to choosing interpretation (a) over interpretation (b).

A new approach that we have recently developed attempts to use all three metrics viz., regularity, planarity and compactness to recover perceptually 'correct' 3D structures from 2D line drawings.

A NEW 3D SHAPE RECOVERY ALGORITHM:

We mentioned earlier that the two features that best characterise a 3D shape recovery process are :

1. the objective function, and
2. the search space traversal strategy.

We begin the discussion of our 3D shape recovery algorithm with a description of these two features.

The objective function:

We ideally wish to obtain that 3D configuration which, while having planar faces is maximally regular. Furthermore, the recovered 3D structure should be compact, i.e., it should not be necessary to invoke the possibility of excessive foreshortening on any line segment to relate the observed projection with the 3D structure.

The search space traversal startegy:

One of the simplest techniques to minimize a cost function is gradient descent. In the present case, we could define a cost function f as a weighted sum of the three metrics of regularity, planarity and compactness

$$f = w_1\text{*Regularity} + w_2\text{*Planarity} + w_3\text{*Compactness}$$

and then attempt to do a gradient descent on f. Sounds quite reasonable. The fly in the ointment, of course, is how to choose the appropriate w_i's. The relative weights that might be appropriate for one case might be inappropriate for another. The choices of the weights can, at best, be *ad hoc*. Even if one uses sophisticated techniques like the Continuation method [Leclerc, 1989; Fischler & Leclerc, 1992], the values of many parameters have to be chosen in an *ad hoc* fashion.

An even more fundamental problem associated with this approach is illustrated in figure 5. Consider a two-metric optimization situation which requires determining the best value of one metric corresponding to an optimal value for the other. (In the 3D shape recovery context, this is analogous to finding the most regular configuration amongst all possible planar-faceted ones.) The desired metric pair is shown highlighted in the figure.

If the two metrics had been combined into one composite cost function, however, the global minimum might have occured at a very different place, as shown in the lower curve of figure 5. Thus, the combination of the two metrics is likely to shift the optimal solution from its 'correct' position. This argument has also been mentioned in the context of regularization. By the inclusion of the regularizing term, one might end up with a solution quite unrelated to the original problem.

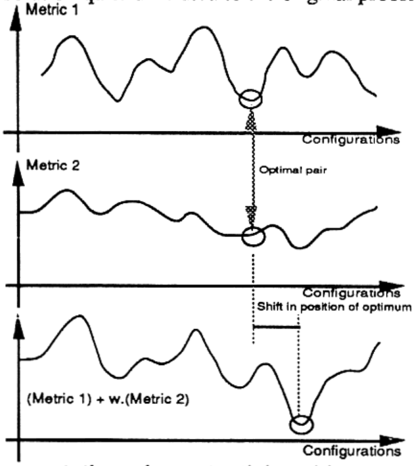

Figure 5. A schematic representation of a potential problem associated with composite cost functions. For some optimization scenarios, such combinations of metrics might result in shifting the optimal solution from its 'correct' position. See text for details.

Is there some way around using a parameterized cost function? Can the need for *ad hoc* choices be reduced? Can the same algorithm work on several different cases without requiring parameter tweaking for each individual case? A tall order, undoubtedly, but it seems that a space traversal strategy that answers these questions affirmatively does exist. We next describe this strategy.

To motivate the particular search space traversal technique we have used, let us begin with a simple thought experiment. Imagine that one is given a 2D line drawing that one wishes to derive the 3D shape of in accordance with the objective function decribed previously. How might one get to the desired shape? In other words, what kinds of intermediate stages should one expect to pass through? One natural way of getting to the desired shape is to incrementally modify the originally planar configuration so that at every intermediate step one obtains the most regular locally planar configuration (each facet planar). This can be thought of as doing gradient descent in regularity space where the points considered in the space correspond to the different planar 3D configurations. The first local minima reached in this fashion is reported as the recovered 3D shape.

This, then, is the basic idea: get to the final configuration via a series of intermediate maximally regular planar configurations. (A reader familiar with computer algorithms might notice the similarity of this approach to a 'greedy' strategy - achieving a globally optimal solution through locally optimal choices.) The compactness heuristic may be implemented on top of this progression; the process can be terminated whenever the (lack of) compactness of a recovered configuration exceeds a preset threshold. This threshold is arbitrary and is the only *ad hoc* feature in the algorithm. But, this *ad hoc*ness is quite acceptable since the choice of the threshold is uniform over all objects. Consequently, no parameter tweaking is required to run the algorithm on different cases.

That was an informal, qualitative description of the kind of search space traversal strategy one might adopt. We now describe precisely how to implement this strategy for a beads-on-wires situation. To begin with, we state a proposition and two corollaries that will be made use of for justifying and explaining the behavior of the particular search space traversal strategy we adopt.

Proposition 1:

For any two lines with a common end-point, their included angle varies monotonically from θ_0 to θ_1 (where θ_0 and θ_1 are respectively, the included angles when the plane defined by the two lines is parallel to the image plane (p=0, q=0) and when it is oriented such that p = p_f and q = q_f for some arbitrary p_f and q_f) for surface orientations (k p_f, k q_f) as k varies from 0 to 1. In other words, the included angle varies monotonically in going along the straight line joining the origin with an arbitrary point in gradient space.

Proof:

Consider three arbitrarily chosen points A, B, and C as shown in figure 6(a). Without loss of generality, we may assume that $(x_3, y_3, z_3) = (0, 0, 0)$. The vectors $\overline{AB}$ and $\overline{CB}$, then, are (x_1, y_1, z_1) and (x_2, y_2, z_2) respectively. The included angle $\angle$ABC may be obtained by computing the dot-product of these two vectors.

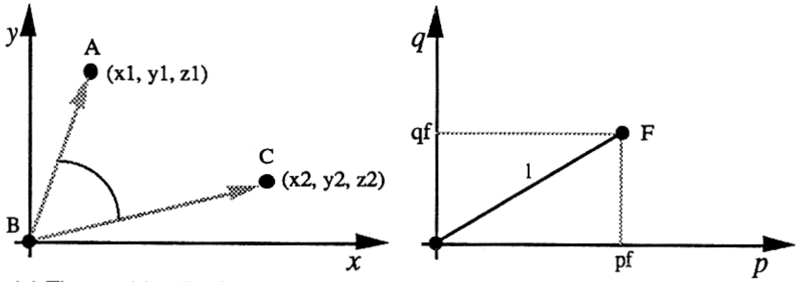

Figure 6. (a) Three arbitrarily chosen points in x-y space. (b) An arbitrarily chosen point in gradient space. See text for details.

Let F = (p_f, q_f) be an arbitrary point in gradient space (see figure 6(b)). We wish to show that as we vary the orientation of the surface ABC so as to move along the line 'l' from the origin (p = 0, q = 0) to point F (p = p_f, q = q_f) in gradient space, the included angle varies monotonically with the distance moved along the line. We can parametrize this distance by a multiplicative factor k $\in$ [0, 1] such that (k p_f, k q_f) defines the intermediate surface orientations along 'l'. The task then is to show that angle $\angle$ABC varies monotonically with k.

Taking the dot-product of $\overline{AB}$ and $\overline{CB}$, we have

$$\overline{AB} \cdot \overline{CB} = |\overline{AB}|\,|\overline{CB}| \cos\theta$$
$$= x_1 x_2 + y_1 y_2 + z_1 z_2$$
$$\therefore \cos\theta = (x_1 x_2 + y_1 y_2 + z_1 z_2)/|\overline{AB}|\,|\overline{CB}|$$
$$= \frac{x_1x_2 + y_1y_2 + z_1z_2}{\sqrt{x_1^2 + y_1^2 + z_1^2}\sqrt{x_2^2 + y_2^2 + z_2^2}} \qquad \text{.........(1)}$$

By definition of surface depth gradients,

$$z_i = kp_f x_i + kq_f y_i$$
$$= k(p_f x_i + q_f y_i)$$
$$= kz_{fi}$$

where z_{fi} is the depth of the point i when surface ABC has depth gradients equal to p_f and q_f. Substituting $z_i = kz_{fi}$ into (1), we have

$$Cos\theta = \frac{x_1x_2 + y_1y_2 + k^2 z_{f1} z_{f2}}{\sqrt{x_1^2 + y_1^2 + k^2 z_{f1}^2}\sqrt{x_2^2 + y_2^2 + k^2 z_{f2}^2}}$$

$$\therefore \qquad \theta = Cos^{-1}\frac{x_1x_2 + y_1y_2 + k^2 z_{f1} z_{f2}}{\sqrt{x_1^2 + y_1^2 + k^2 z_{f1}^2}\sqrt{x_2^2 + y_2^2 + k^2 z_{f2}^2}}$$

To demonstrate that θ varies monotonically with k, all that is required is to show that the partial derivative of θ with respect to k i.e. $\partial\theta / \partial k$ is either positive or negative but not zero. Some conceptually simple but tediously long algebraic manipulations prove that this indeed is the case. $\partial\theta / \partial k$ is zero only if the three points A, B and C are collinear. For any other configuration of the three points, $\partial\theta / \partial k$ is either positive or negative, showing that θ does indeed increase or decrease monotonically with k.

Corollary 1:

For regular polygons, as one progresses from a non-regular projection (parallel to the image plane and therefore at the origin of gradient space) to a surface orientation α that renders the polygon maximally regular, the standard deviation of all included angles (SDA) decreases monotonically.

Proof:

This is a direct outcome of the result that all included angles in the polygon converge monotonically to the same value as the surface approaches the orientation α.

Corollary 2:

In gradient space, the straight line joining the point nearest to the origin which represents an orientation that renders a polygon maximally regular defines the path of steepest descent for SDA values.

Proof:

The truth of this assertion follows straightforwardly from proposition 1 and corollary 1. To aid intuition, observe that the SDA decreases *monotonically* by the *same amount* in going from the origin to any point in gradient space that represents an orientation of maximum regularity. The rate of descent is, therefore, inversely proportional to the distance of the point from the origin.

With some formal groundwork in place, we proceed to describe our search space traversal strategy.

Consider figure 7. It schematically shows two independent cost spaces - one for regularity and the other for planarity. Any configuration of the beads-on-wires setup corresponds to specific locations in these two spaces. When the algorithm begins off, the deviation from planarity is, of course, 0 since the input line-drawing is planar. Since each of the n vertices can be moved up or down independently by the chosen step-size and one such movement constitutes a 'step' of the algorithm, we have 2n possible successor states from any given state of the beads configuration. These choices correspond to having 2n possible ways of moving in the two spaces (shown by arrows in the figure).

For regular polygons, we proved in corollary 1 that the desired configuration that maximizes regularity and compactness can be reached from the origin by following the path of steepest descent. For a single polygon, then, we would like to choose those of the 2n possible options which decrease the SDA. From amongst the multiple candidates that decrease the SDA, we choose those that minimize the deviation from planarity i.e. which minimize the up-climb in non-planarity space. In the situation, where there are more than one such possible options, we choose the one that decreases the SDA the most. For a multifaceted object, the same approach applies since the overall SDA is just the sum of the SDA's of the individual faces. This completes the description of the search space traversal strategy.

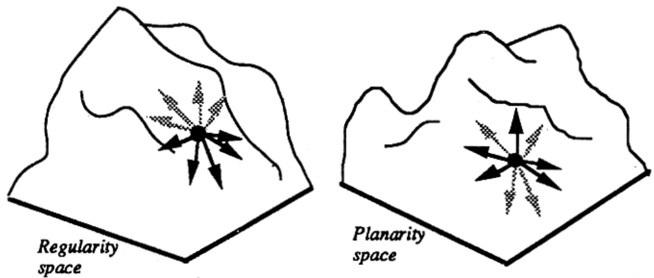

Figure 7. Our approach involves keeping the regularity and planarity cost spaces separate and having the latter constrain the choices in the former. See text for details.

The decision to choose the configuration that minimizes the up-movement in non-planarity space has precisely the effect that we want. It attempts to bring the whole bead configuration to a maximally planar state as fast as possible. Thus, though individual bead movements introduce little bits of non-planarity (imagine tweaking just one bead of a square), the subsequent bead movements are so chosen as to get the whole setup back into a planar state (but one that is more regular than the previous one). We expect the non-planarity plot to show little bumps indicating how the recovered structure goes from one planar configuration to another through slightly non-planar intermediate states. Happily enough, this is precisely what happens and in figure 8 we reproduce two plots of non-planarity versus time for showing the behavior of the algorithm as it recovers the 3D shapes corresponding to two different line drawings.

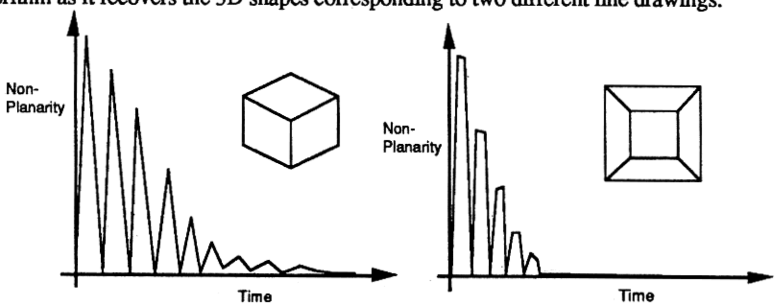

Figure 8. The non-planarity versus time plots for two sample objects demonstrate that the algotithm systematically examines maximally regular planar faceted configurations.

One final point remains to be mentioned. How do we know that the intermediate 3D configurations explored (with planar facets) will be the 'right' ones? For instance, in figure 9, what prevents us form having 1' as an intermediate state between 0 (the initial configuration) and 2 (the desired final state) rather than 1? The answer to this question makes explicit the inherent assumption embodied in the algorithm and, therefore, also serves to highlight its limitations. The assumption is that the final orientation of each face is the most compact and maximally regular one. Now, as was shown in corollary 2, the straight line joining the origin of gradient space to this point also defines the direction of steepest descent in regularity. Performing steepest descent on the sum of the individual SDA metrics therefore constrains the intermediate configurations to be *1D interpolants between the origin and the final configuration.* This is what rules out exploring state 1' instead of 1 in figure 9.

What does this discussion say about the limitation of the algorithm? Simply put, it is this - if the perceptually correct interpretation does not have all faces as being maximally regular and compact, then our attempt to get at such a solution using the algorithm (and even actually finding it) is bound to produce 'incorrect' results (i.e. not in conformity with human perception).

Figures 10 (a) and (b) presents two examples showing the 3D structures recovered by our approach from the input 2D line drawings.

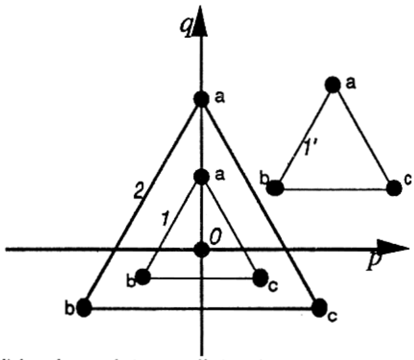

Figure 9. Of all the possible planar intermediate stages, our approaches chooses the one which maximizes the descent in regularity space.

In practice, the algorithm can tolerate a substantial amount of slack in terms of deviations from planarity and regularity in the faces of the intermediate and final configurations. So long as a majority of the faces are well behaved, the erratic behavior of the rest is overshadowed. More precisely, so long as the well-behaved nature of a majority of faces ensures our following the right descent path in the cost space, slack on a few of the faces does not matter. Of course, the slack can be distributed across all faces instead of being confined to just a few. Then the amount of slack permissible per face is proportionately reduced. Figure 10(c) shows a line drawing which is correctly interpreted by our algorithm as an approximate cube with slightly warped faces. The result shows that a significant amount of deviation from planarity can be tolerated by the approach.

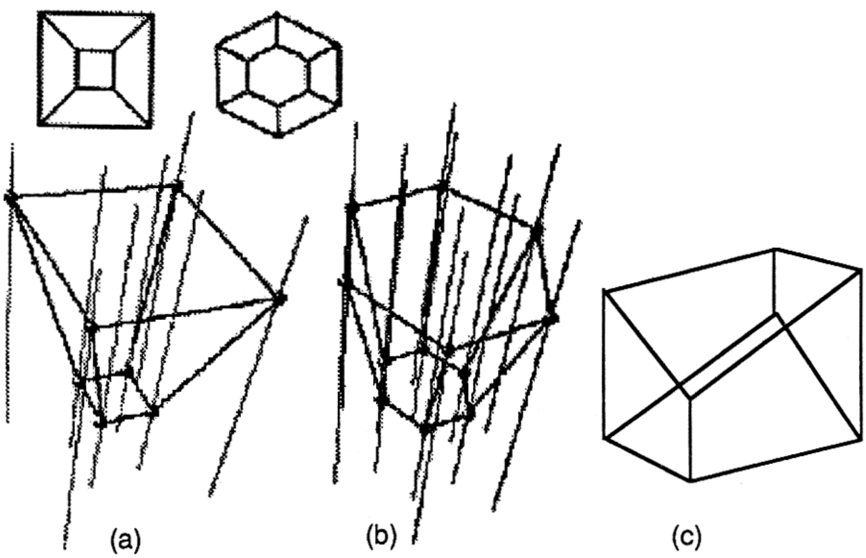

Figure 10. (a), (b) Two examples of 3-D shape recovery using constraints of symmetry, planarity and compactness. (c) A line-drawing that is 'correctly' interpreted by our algorithm as an approximate cube with slightly warped faces.

CONCLUSION:

We have proposed a reasonably robust method for recovering 3-D wireframes from 2-D line drawings of simple polyhedral objects. Besides serving as a computational model of a perceptual process, this method also holds promise as a means of simplifying man-machine interaction. In CAD/CAM for instance, a part could be described using only its 2D projection; the 3D reconstruction could be accomplished automatically by the machine. Such an ability would also be of crucial significance for any image understanding system. A word of caution though. We do not wish to suggest that the ideas for shape recovery we have just described constitute a general solution for the problem of interpretation of arbitrary line drawings; while some of these ideas may capture general truths, as currently implemented they can be expected to work only in restricted domains. In particular, the current system cannot handle smoothly curved objects and scenes involving occlusion. Furthermore, for arbitrary line drawings, the rather non-specific notions of symmetry, planarity and compactness, among many others, are confounded with constraints specific to particular objects imposed by high level knowledge. Trying to obtain the completely general solution is an endeavor that is equivalent to trying to figure out how the brain works. This is probably a trifle beyond the scope of this paper.

References:

Attneave, F., Frost, R. 1969. The determination of perceived tridimensional orientation by minimum criteria. *Perception and Psych.* 6:391-396.

Barrow, H. G., Tenenbaum, J. M. 1981. Interpreting line-drawings as three-dimensional surfaces. *Artificial Intelligence,* Vol. 17, Nos. 1-3, pp. 75-116.

Brady, M., Yuille, A. 1983. An extremum principle for shape from contour. *MIT AI Lab Memo* 711.

Clowes, M. B. 1971. On Seeing Things. *Artificial Intelligence,* Vol. 2, No. 1, pp. 79-116.

Fischler, M. A., Leclerc, Y. G. 1992. Recovering 3-D wire frames from line drawings. *Proceedings of the Image Understanding Workshop.*

Horn, B. K. P. 1975. Obtaining shape from shading information. In *The Psychology of Computer Vision,* ed. P. H. Winston New York: McGraw-Hill Book Co.

Huffman, D. A. 1971. Impossible objects as non-sense sentences. In *Machine Intelligence 6* eds. Meltzer and Michie pp. 295-323, Edinburgh Univ. Press.

Kanade, T. 1981. Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence,* Vol. 17, Nos. 1-3, pp. 409-460.

Koffka, K. 1935. *Principles of Gestalt Psychology.* New York: Harcourt Brace.

Leclerc, Y., (1989). *Constructing Simple Stable Descriptions for Image Partitioning.* Intl. J. Comp. Vision 1, 73-102.

Mackworth, A. K. 1973. Interpreting pictures of polyhedral scenes. *Artificial Intelligence,* Vol. 4, No. 2, pp. 121-137.

Marill, T. 1991. Emulating the human interpretation of line-drawings as three-dimensional objects. *Intl. J. of Comp. Vis.* 6:2 pp. 147-161.

Roberts, L. G. 1965. Machine perception of three-dimensional solids, in J. T. Tippett et al. Eds., *Optical and Electro-Optical Information Processing,* MIT Press, Cambridge, MA.

Waltz, D. L. 1972. Generating semantic descriptions from drawings of scenes with shadows. In *The Psychology of Computer Vision,* ed. P. H. Winston New York: McGraw-Hill Book Co.

Witkin, A. P. 1981. Recovering surface shape and orientation from texture. *Artificial Intelligence,* Vol. 17, Nos. 1-3, pp. 17-45.

Yuille, A. L. 1987. Shape from shading, occlusion and texture, *Memo* 885, MIT AI Lab.