Recovering Partial 3D Wire Frames Descriptions from Stereo Data

Stephen B Pollard, John Porrill and John E W Mayhew
AIVRU
University of Sheffield

We describe the design of modules in the current version of the TINA stereo based 3D vision system responsible for the recovery of geometric descriptions and their subsequent integration into partial wire frame models. The approach differs considerably from that described in an earlier version of the TINA system [2, 3]. Our strategy for the construction of partial wire frame descriptions is broken into the following three processing stages:

(1) *Stereo processing. This is edge based and uses camera calibration information to provide a restriction on allowable matches between edges extracted from the left and right images. The algorithm does not, however, at this stage recover an accurate estimate of disparity of each pair of matched edges.*

(2) *Geometrical description. This relies on the identification of higher level 2D primitives from the edge strings in one, other or both images and their subsequent combination with stereo matching, and calibration data to allow the recovery of 3D primitives.*

(3) *Wire frame generation. Based upon 2D and 3D proximity and connectivity heuristics potential wire frame components are identified.*

1. Introduction

Feature based stereo (either binocular or trinocular) provides one of the simplest, best understood and most robust methods for the passive recovery of three dimensional information from image data. Furthermore, given the availability of suitable *a priori* information concerning the structure of the imaging system, geometrical descriptions recovered from stereo are metrical if (depending upon the care taken in their recovery) not always accurate. However, for many robotic tasks the availability of unstructured depth data alone is not sufficient for their effective completion. In general the data must be structured in some way to be of use.

This paper is concerned with the interpretation of edge based stereo data with a view to the construction of partial 3D wire-frames. The latter is a description that makes explicit potential wire frame primitives and hypothesised connections between them. It does not however involve the completion of wire frames to the status of boundary descriptions of physically realisable objects. As such partial wire frames are allowed to include descriptions that arise from changes in surface reflectance properties whilst full wire frames are not. We conjecture that full wire frame descriptions can only be reliably recovered in the presence of domain knowledge. Partial wire frames on the other hand are built purely bottom up using only local heuristics.

Of course partial wire frames descriptors will of themselves be of considerable use. For example we shall show that Geomstat [1] a software package that uses statistical estimation theory to enforce geometrical constraints (such as intersection, parallelism, orthogonality, continuation, coplanarity, etc) can be effectively driven over partial wire-frame data.

2. Edge Based Stereo

Edges are obtained, to sub pixel acuity, from grey-level images by a single scale high frequency application of the **Canny** edge operator [4]. The effect of the operator is essentially to identify contours corresponding to above threshold maxima of image intensity gradients. Noise suppression and scale selection are achieved by a pre-smoothing stage involving convolution with a gaussian mask (the size of the mask determining the degree of smoothing). The high frequency operator used here employs a gaussian mask of sigma 1.0 (diameter of 9 pixels when thresholded at 0.001 of the peak value). Fortunately the two dimensional gaussian smoothing can be achieved through two 1 dimensional convolutions (eg. first along the rows and then the columns). Further improvements in efficiency can be achieved through the use of recursive filters [4, 5].

Recent and more effective versions of our stereo algorithm, called PMF, are used to match corresponding edges [6]. Essentially matches between edges from the left and right images are preferred if they mutually support each other through a disparity gradient constraint (ie. that the ratio of the difference in their

disparity to their physical separation is below a threshold of 0.5) and if they satisfy a number of higher level grouping constraints, most notably uniqueness, ordering (along epipolars) and figural continuity. Initially individual matches are restricted to be of similar contrast (within a factor of 3) and similar orientations (consistent with a disparity gradient limit; as a consequence of which edges perpendicular to the epipolar direction are permitted to reorientate more extremely than those runing along it).

Using information about the geometry of the camera's obtained from a calibration stage allows the matching problem to be reduced to a one dimensional search along corresponding epipolars. Furthermore if the edge maps are transformed to an alternative parallel camera geometry that is equivalent to the original, the epipolar constraint becomes a same raster constraint with obvious benefit in reducing the complexity of the stereo matching algorithm.

The results of this stage of processing are connected strings of edge primitives from the left image some of which have associated with them edge primitives with which they are thought to match in the right image. It is possible for connected edges to run along the epipolar (raster) hence (especially if the edge direction is almost horizontal) it is sometimes difficult (and no attempt is made here) to resolve matches uniquely at this stage.

3. Obtaining Geometrical descriptions

While the availability of suitable estimates of the epipolar geometry and/or the intrinsic and extrinsic camera parameters is fundamental to the accurate recovery of qualitative and quantitative shape descriptors from stereo; practical, robust and generic visual capabilities require much attention to the method used to incorporate disparity information in the construction of the three dimensional descriptors themselves.

We begin this section by considering three strategies by which stereo edge data can be combined (assuming known epipolar geometry) to recover 3D scene primitives. In order to clarify the issues involved we place each method in the context of recovering a plane curve from a string of matched edge data (see also figure 1).

The scheme used in the early versions of TINA [7] is to begin by obtaining estimates in a disparity space space (chosen to be the unbiased space [Xl, Xr, Y]) from the sub-pixel estimations of individual edge correspondences (taking account of their off-epipolar differences). Subsequently (and after suitable segmentation) a plane is fitted to the disparity data by orthogonal regression. Unfortunately this process is unable

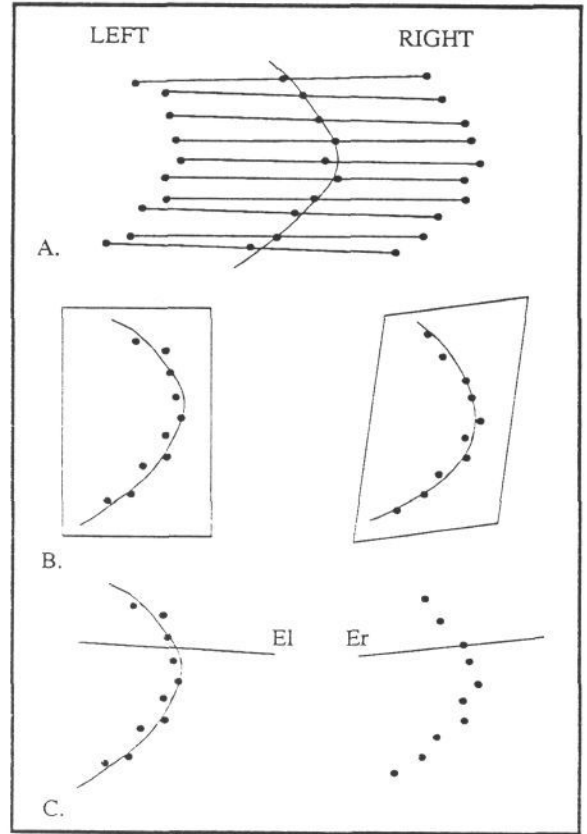to recover from the initial sub-optimal estimate of the individual disparity values.

Figure 1.

The figure illustrates three possible strategies for the combination of stereo matching data to recover plane curve descriptors. In each sub pixel edge locations are shown in the left and right image as dots. In A left and right image primitives are combined and individual disparity estimates obtained prior to fitting. In B optimal fitting is first performed in the left image with a subsequent recovery of the affine transform required to minimise the sum squared error of right hand edge points to the transformed curve. In C (the preferred method) consider the epipolar that passes though a point in the right image and the corresponding epipolar in the left (labeled **el** *and* **er** *respectively). In general no left edgel will not lie exactly on* **el** *and hence the disparity measure will be suboptimal. The edge comprised of the edgels in the left image can be approximated by the curve C (if possible and computationally plausible, optimally). The optimal estimate of disparity is obtained from the intersection of the epipolar* **el** *with C.*

In the method proposed by Cooper, Hung and Taubin [8], edges are initially brought into correspondence at the level of extended edge strings. Given surface and curve parameterisation schemes (in their case plane and polynomials respectively) maximum-likelihood estimation is employed to [implicitly] refine match selection to give optimal estimates of both the surface and curve parameters. In practice curve estimation in one or other image precedes the stage of surface estimation and the parameterisation of the latter is in terms of an affine transforms between the images (whilst in theory it could, the affine transform is not restricted to conform to the known epipolar geometry). The primary disadvantage of this approach (beyond purely computational considerations) is its lack of uniformity with respect to possible curve and surface parameterisations.

Here we exploit a more pragmatic, though not optimal, solution to the problem which lies between those described above. As with Cooper et al's method high level feature grouping (edge strings, straight lines, conic sections, cubic splines, segmented regions, etc) is performed prior to disparity detection. Optimal estimates of disparity are obtained along the 2D geometrical descriptors for each matched edge point from the intersection of the descriptor with the epipolar corresponding to the sub-pixel location of the matched point. The optimal disparity estimates are then combined in a second stage of fitting (in this case to a plane) in 2D arc position against disparity. Hence we assume only the availability of a generic edge matching capability that is able to bring into rough correspondence edge primitives from the left and right images.

In addition to its uniformity over curve and surface descriptions (which are illustrated in the next section) there are a number of additional advantages to this approach over either or both of the above:

(1)   The 2D descriptions are reasonably robust.

(2)   The 2D descriptions can be obtained economically.

(3)   Edge matching is reasonably generic.

(4)   Disparity to segment fitting is optimal.

(5)   Geometrical constraints (eg. planarity) can be exploited easily.

(6)   The 2D descriptions remain available for unmatched features.

3.1. Describing Edge Strings

In this section we describe methods for the segmentation, recovery and description of various 2 and 3D primitives.

The algorithm to find straight lines, for example, uses a recursive fit and segment strategy (see figure 2). Segmentation points are included when the underlying edge string deviates from the current line fit. The actual fit (first in the image and then in arc length against disparity) is computed by orthogonal regression. Notice that it is simple matter to eliminate occasional erroneous matches if their disparity values are inconsistent with the majority. A good example of performance upon objects for which straight line approximation is appropriate is given in figure 6 parts (a) and (b).
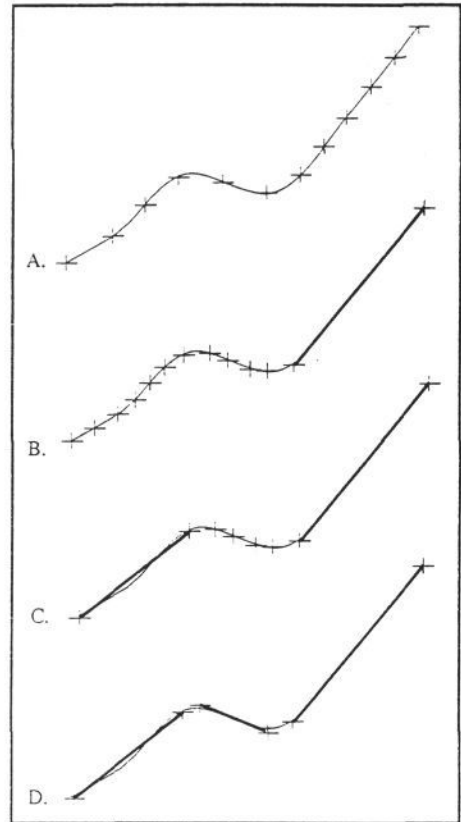

Figure 2.

Polygonal approximation by recursive fit and segmentation. In A the initial sampling of an edge string is shown. The longest linear section is found by efficient heuristic search and fitted by orthogonal regressions and is shown in B. Further recursion at this sampling scale recovers no further significant straight sections and the remaining string this is further sub sampled (also shown in B). Continued recursion at the new scale recovers further straight sections shown in C and D. This method has a number of advantages over traditional polygonal approximation and is easily extended to other primitive types.

Elliptical sections and their confidence envelopes are obtained optimally using a bias corrected Kalman filter [9]. However a reasonably large quadrant of the ellipse is required for accurate description. A sub optimal (but acceptable) estimate for the ellipse in 3D is obtained through a weighted least square plane fit through the disparity data. Weights are applied according to the orientation of the ellipse at the point at which the disparity measure is obtained. In general orientations closer to the epipolar direction are subject to greater pixelation and epipolar calibration error than those orthogonal to it.

Generic plane curves are be treated in a fashion analogous to that for ellipses. The principal issues are those of representation, segmentation and combination of the 2D and 3D curve descriptors. The desired representation should balance the competing properties of economy, power and uniformity (between the descriptions in the 2D image, disparity space and the 3D scene). The results presented below were obtained using standard cubic splines.
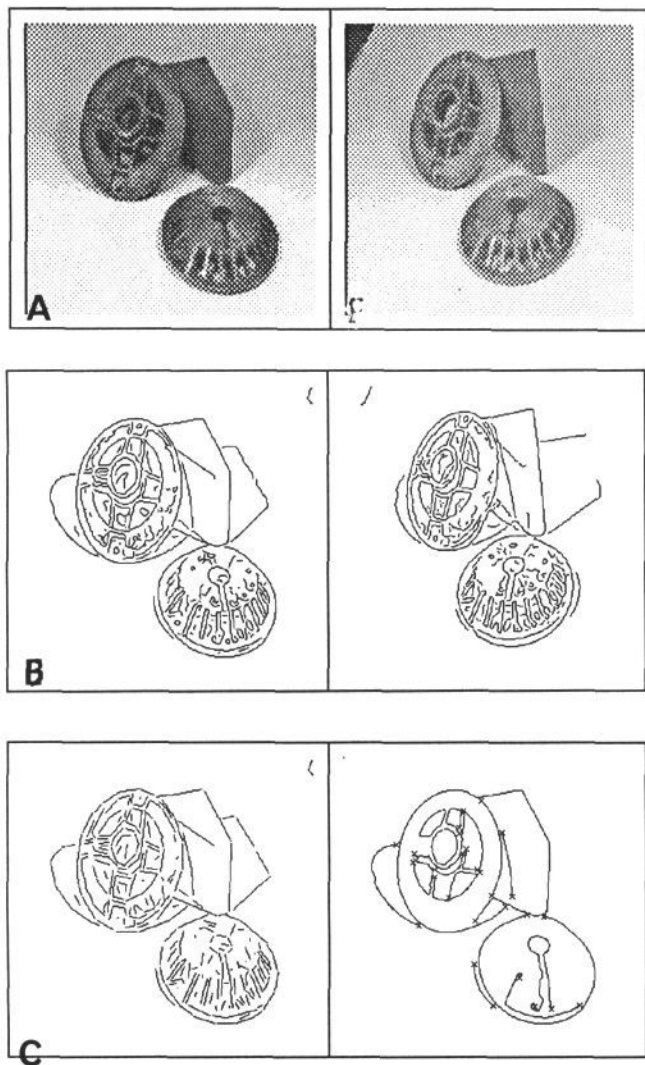

Figure 3.

A is a stereo image pair [512 pixels square] of a triple of machined parts. In B are the canny edge points identified in each image, and in C are respectively on the left and right hand sides a polygonal approximation and a few extended edge strings obtained from the left image.
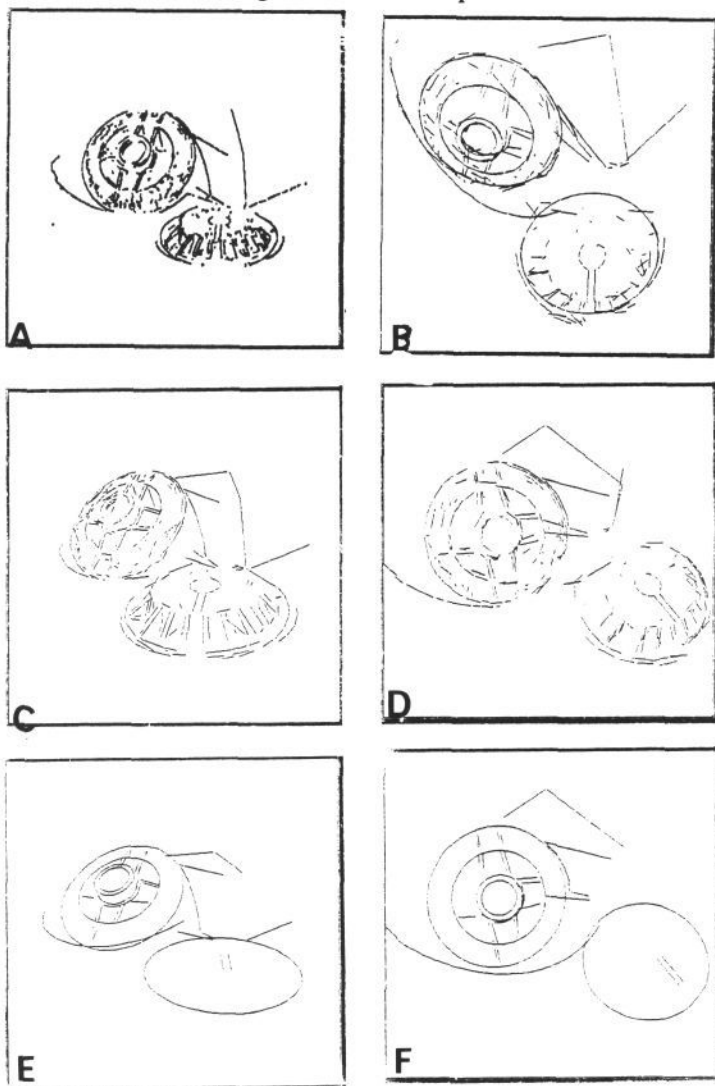

Figure 4.

A shows depth data to approximate sub pixel location for uniquely matched (1 to 1) edge points. In B straight line, and selected conic and planer curve descriptions of the scene are viewed from above. C and D show just the 3D straight line approximations from two other viewpoints. Whilst straight line sections provide some form of description they are fragmented and noisy for these kinds of object. E and F show curve sections from the same viewpoints with a few straight sections included to provide context.

Examples of straight line, conic and plane curves obtained from machined objects are given in figures 3, 4 and 5.

We do not advocate that simple unique descriptors can be recovered for all edge strings on the basis of bottom up processing alone. It may be sensible to recover multiple descriptions of the same data for edge sections that are close to threshold boundaries in terms of their chosen description, for example a space curve may also be described as a sequence of plane curves or even its polygonal approximation. The appropriate descriptor may only become apparent at higher levels in the processing hierarchy.
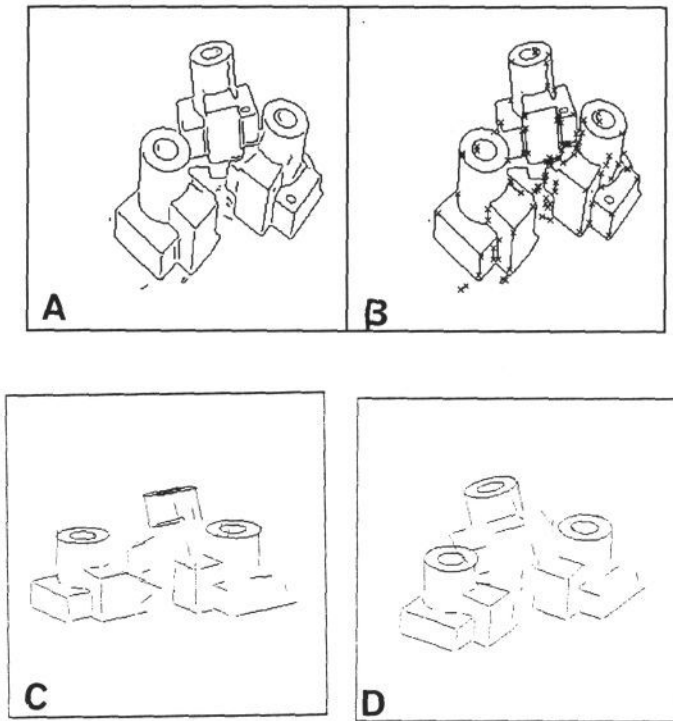

Figure 5.

3 widgets. Edge data and connected strings in A and B respectively. Two view points of the recovered 3D data (straights and ellipses) in C and D.

4. Building Partial Wire Frames

We have also found the generation of partial wire frame descriptions to be easier, computationally cheaper and more robust when based upon the image projection of an object rather than the three dimensional description of the identified image primitives alone. Consider the object in figure 6. We currently limit wire frame descriptors to linear sections alone (though the vocabulary of the system that constructs wire frames is designed to be easily extended to include other types of primitive). It is straight forward

to choose simple neighbourhood criteria in the 2D image as it is less vulnerable to noise and feature drop out; these form the basis of potential connectivities. Each line section is treated as a triple which includes the line itself and its two end points, this makes it straight forward to differentiate connections between the ends of lines (potential vertices, etc) and other forms of connection (parallelism, tee junctions, etc). A few simple heuristics in the image data can be used to identify various vertex and connectivity (eg. collinearity) types. Vertices in the figure were recovered using the following simple strategy:

(1) Identify all potential vertex connectors (PVC) between the ends of line sections. That is: within a local neighbourhood; close to point of line intersection; closer than the other end of the same line to the point of intersection.

(2) Consider isolated connectivity graphs amongst PVC's.

(3) Find the computationally trivial simple pairwise vertices first. That is connectivity graphs that consist of a single PVC.

(4) Now find high order vertices (trihedral vertices and above). These can only arise from mutually consistent cliques of cardinality at least 3 identified in the graph structure. Additional consistency in terms of the actual location of the vertex is also enforced (that is the location of the intersection points identified between all constituent line pairs must be in general agreement). PVC's included in such vertices are removed from their graphs (this may cause some extra fragmentation, hence stage 3 is repeated on them).

(5) Finally, identify the best of the remaining pairwise vertices. Where best is defined as the minimum sum of distances from the end points to the point of intersection. Each end point is allowed to occur in only a single such vertex.

Vertices thus identified (see the figure) can be interpreted amongst the related 3D primitives (if they exist). Furthermore extra 3D primitives can be hypothesised where their existence is implied by the 2D partial wire frame (for example a 3D line can be generated between a pair of 3D vertices even though it may only exist in the 2D description itself).

5. Conclusions and Future Directions

We have described those components of the current versions of TINA system responsible for the recovery of partial wire frame descriptions. Whilst the work in is far from complete, it does identity a direction in

which to go in order to recover good quality wire frame geometrical descriptions from stereo data. Much work is required in the 2D image to identify meaningful events in the scene in order to apply appropriate constraints to the interpretation of the three dimensional descriptions.

As discussed above we are in the process of combing the bottom up knowledge free strategies presented here with a top down domain-dependent system responsible for inferring completed wire frame descriptions. Domain knowledge for generic object classes (houses are the classic example) is represented as frames. The completion process is then cast as a matching problem and slot filling task of the database against the partial wire frame description. This work is as yet too preliminary to discuss in any greater detail.

Figure 6.

A polygonal scene suitable for primitive partial wire frame description. A and B are the recovered 2 and 3 dimensional straight line approximations respectively. Notice the less complete and more fragmented nature of the 3D data. Connectors implying potential vertex locations recovered on the basis of 2D neighbourhoods are shown in C and D. Cleaned vertices (by Geomstat) found by the method given in the text are shown in E and F. The interpretation of the 2D vertices amongst the 3D data allows extra 3D primitives to be hypothesised.

References

1. Porrill J, Pollard S B, and Mayhew J E W (1987) The optimal combination of multiple sensors including stereo vision, and Image and Vision Computing, Vol 2, No 5 174-180.

2. Porrill J, Pollard S B, Pridmore T P, Bowen J, Mayhew J E W, and Frisby J P (1987) TINA: The Sheffield AIVRU vision system, IJCAI 9, Milan 1138-1144.

3. Pollard S B, Pridmore T P, Porrill J, Mayhew J E W, and Frisby J P (1989) Geometrical modelling from multiple stereo views, International Journal of Robotics Research, vol 8, No 4, 3-32.

4. Canny J. F. (1983) Finding edges and lines in images, MIT AI Lab. Memo 720.

5. Deriche R (1990) Fast algorithms for low-level vision, PAMI 12 No 1, 78-91.

6. Pollard S B, Mayhew J E W and Frisby J P (1985) PMF: A stereo correspondence algorithm using a disparity gradient limit, Perception 14, 449-470.

7. Pridmore T.P., J. Porrill and J.E.W. Mayhew (1987), Segmentation and description of binocularly viewed contours, Image and Vision Computing, Vol 5, No 2, 132-138.

8. Cooper Hung and Taubin (1989) PAMI.

9. Porrill J. (1990) Fitting ellipses and predicting confidence envelopes using a bias corrected Kalman filter, Image and Visi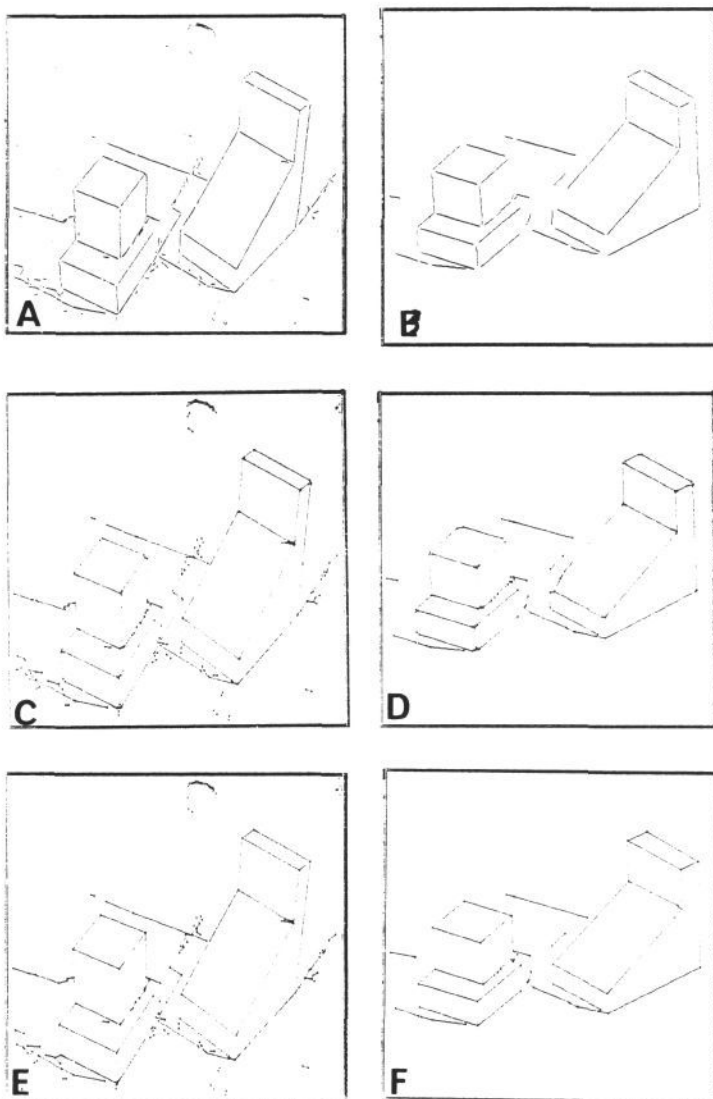on Computing, Vol 8, No 1, 37-41.