THREE-DIMENSIONAL DESCRIPTION OF OBJECTS AND DYNAMIC SCENE ANALYSIS

J. K. Aggarwal

TR-83-1-20 January 1983

Laboratory for Image and Signal Analysis
Department of Electrical Engineering
The University of Texas
Austin, Texas 78712

This research has been supported by the Air Force Office of Scientific Research under AFOSR Contract F49620-83-K-0013.

THREE-DIMENSIONAL DESCRIPTION OF OBJECTS AND DYNAMIC SCENE ANALYSIS

J. K. Aggarwal

Laboratory for Image and Signal Analysis
Department of Electrical Engineering
The University of Texas
Austin, Texas 78712

ABSTRACT: The derivation of three-dimensional structure from two-dimensional images was pursued as the stereoscopic vision problem and more recently as the structure from motion problem in the context of dynamic scene analysis where a sequence of images is processed to compute the structure as well as the motion of a rigid object. The present paper reviews briefly the stereoscopic vision equations and several formulations of dynamic scene analysis equations. It discusses the underlying assumptions and presents a coherent view of the various solutions to the problem of determining three-dimensional structure from several two-dimensional images. Further, it outlines recent results on the computation of three-dimensional structure from a sequence of images where point correspondence in images may be omitted, however viewpoint must be specified for each of the views. Finally, the paper discusses possible directions for future research.

1.0 INTRODUCTION

There are fundamentally two distinct ways of describing three-dimensional scenes. These descriptions are based on: (i) viewer-centered methods and (ii) object-centered methods. An intensity image is a viewer-centered description of the scene whereas, an edge-vertex description specifying relative distances between adjacent vertices is an object-centered description of the scene. The following incomplete list presents the variety of available descriptions:

Viewer-centered descriptions
- Intensity Images
- Sequences of images including stereo pairs
- Range Images
- Perspective and orthographic line drawings

Object-centered descriptions
- Solid-geometric methods
- Generalized cones
- Medial axis transformation
- Surface descriptions

There are descriptions which, strictly speaking, do not belong to either of the above classes. For example, the tomographic description of a volume of tissue. A more detailed discussion of the methods for the description of three-dimensional objects as well as for the acquisition of data is found in Aggarwal et al [1].

The subject of the present paper is the integrating of information contained in a sequence of intensity images with the twofold objective of obtaining the three-dimensional description and the motion of the object. This should be distinguished from another class of efforts which combine several distinct viewer-centered descriptions to obtain a single description of the scene. For example, one may combine an intensity image and a range image to obtain an edge map for the image (see Gil et al [2]).

The research for obtaining three-dimensional structure was initially pursued as stereoscopic scene analysis (see Duda and Hart [3]). Usually, the scene was stationary whereas the camera moved (or there were two cameras) to obtain the two images. More recently, several researchers have considered the problem of several views. The following sections review the works of Ullman [4], Roach and Aggarwal [5], Nagel [6], and Tsai and Huang [7], after presenting briefly the perspective equations and the stereoscopic vision problem as discussed in [3]. This is followed by a discussion of a method based upon silhouettes developed by Martin and Aggarwal [8] together with recent results. The last section of the paper discusses the possible directions of future research. The present paper emphasizes the three-dimensional character of dynamic scene analysis in contrast to the earlier reviews by Martin and Aggarwal [9,10] where the entire area of dynamic scene analysis was discussed.

2.0 PERSPECTIVE EQUATIONS

For the case of the simple geometry where the image plane is in front of the lens center, the global coordinate system coincides with that of the image plane and the optical axis aligns with the y-axis, the image plane coordinates of the perspective projection of a point in space are related as follows:

$$\frac{X}{F} = \frac{x}{F+y} \qquad Y = 0 \qquad \frac{Z}{F} = \frac{z}{F+y} \tag{1}$$

Here, (x,y,z) is the point in space, (X,0,Z) is its projection on the image plane as shown in Figure 1. It is clear from the geometry of the figure, and the equations of projection (1), that the position of the image point is uniquely determined from the point in space. However, given the image point, one is able to fix only the projecting ray. This becomes evident if the projection equations are rewritten as

$$x = \frac{X(F+y)}{F} = \frac{Xz}{Z}, \tag{2}$$

or better still using a homogeneous coordinate system as

$$\begin{vmatrix} x \\ y \\ z \end{vmatrix} = \frac{F}{F-Y}\begin{vmatrix} X \\ Y \\ Z \end{vmatrix} \tag{3}$$

where Y is a parameter; and for various values of Y, the point (x,y,z) traces out the projecting ray.

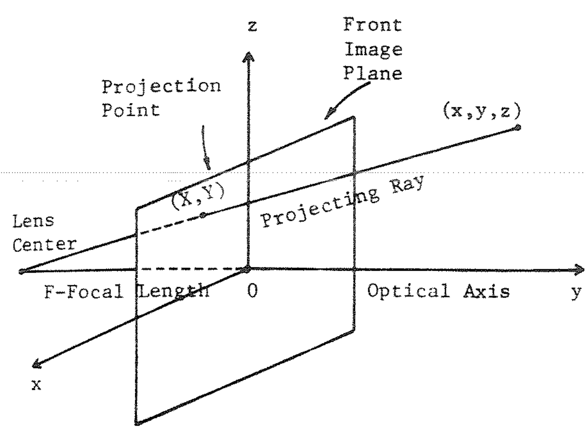

Fig. 1. Lens center and image plane geometry.

For the case where (x_0,y_0,z_0) represents the gimbal center, and the optical axis is panned through an angle θ and tilted through an angle ϕ, and (ℓ_1,ℓ_2+F,ℓ_3) represents the constant offset of the image plane center (relative to the gimbal center), the relationship between the image plane coordinates and spatial coordinates is considerably more complicated as given in [3]. The inverse perspective equations are also given there. The above simpler equations (1),(2), and (3) may easily be derived from the complicated equations by simply assuming $\theta=0=\phi$, $x_0=y_0=z_0=0$ $\ell_1=0=\ell_3$, and $\ell_2=-F$. The derivation of the above equations together with an excellent discussion of perspective equations is presented by Duda and Hart [3].

In order to locate the spatial coordinates of a point in space from its image plane coordinates, one may employ a method called stereoscopy, based upon two image plane views. Each image plane point will give rise to a projecting ray and given that the point lies on both projecting rays, it lies on the intersection of the two projecting rays. This assumes that one has established the correspondence of points in the two images. In its generality, the correspondence problem is difficult and it is not discussed in the following. The interested reader is referred to [11,12].

The stereoscopy arrangement consists of two image planes with two lens centers. Let the lens centers be given by $\vec{L}_1$ and $\vec{L}_2$, and $\Delta=\vec{L}_2-\vec{L}_1$ gives the baseline vector as shown in Figure 2. If $\vec{U}_1$ and $\vec{U}_2$ denote the unit vectors along the two projecting rays then the equation

$$a\vec{U}_1 = \Delta + b\vec{U}_2 \qquad (4)$$

determines the point of intersection of the two projecting rays, and thus the point in space, where a and b are suitable scalar constants. In a noisy environment such a and b will not exist since the two rays may not intersect. The problem may be reformulated as a minimization problem with the penalty function

$$J(a,b) = \| a\vec{U}_1 - (\Delta + b\vec{U}_2)\| \qquad (5)$$

with the minimum being reached for the values of parameters a_0,b_0. The coordinates of the points in space are approximated by

$$\vec{v} = (1/2)(a_0\vec{U}_1 + (\Delta + b_0\vec{U}_2)) + \vec{L}_1, \qquad (6)$$

the midway point of the shortest distance between the two projecting rays. The solution to the above problem is simple and may be found in Duda and Hart [3].

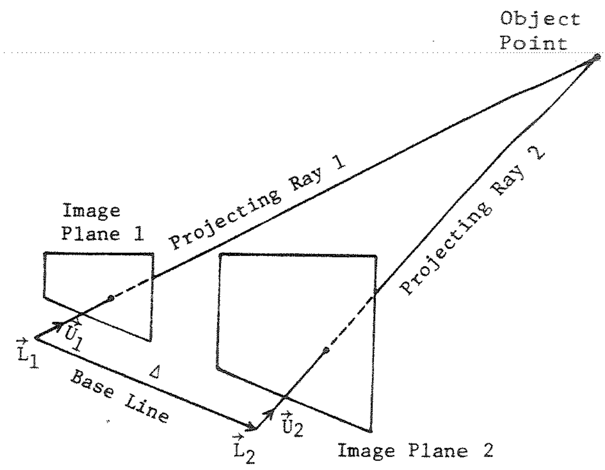

Fig. 2. Stereoscopy arrangement with two image planes.

This early work forms the basis of the more recent work where several researchers have considered the problem of computing the three-dimensional structure from several views together with the motion parameters of the object. The works of Ullman, Roach and Aggarwal, Nagel, and Tsai and Huang are discussed. It may be emphasized that notations used by various authors are different, and at times they are considering slightly different problems. The purpose of the review is to present a coherent view of the problem and its solutions.

3.0 EQUATIONS FOR DYNAMIC SCENE ANALYSIS

Several researchers have considered the equations necessary for dynamic scene analysis of images obtained by perspective transformations of scenes containing three-dimensional objects. The present section reviews the works of Ullman [4], Roach and Aggarwal [5], Nagel [6], and Tsai and Huang [7]. The underlying assumption in all these works are: (i) the objects are rigid and (ii) the correspondence of points between images has already been established.

In the above presentation of section 2 the y-axis is used as the optical axis. However, this convention is not followed universally. In particular, for the works discussed in the present section, the original notation of the author is used. For example, in sections 3.2 and 3.4, the z-axis is the optical axis, whereas in 3.1 and 3.3 the y-axis is the optical axis. The reader is forewarned of the obvious inconsistency. In addition, it may be noted that lower case triplets like (x,y,z) denote points in 3D space in terms of the global coordinate system, whereas upper case pairs (X,Z) or (X,Y) denote image plane points in the image plane coordinate system.

3.1 Polar Equations

Ullman considers the constraints imposed by the perspective projection and combines them with the constraints imposed by motion of a rigid object to derive the structure of the object. Let (x_i, y_i, z_i), $i=1,2,3$ represent the spatial coordinates of points on an object undergoing a translation $(\Delta x, \Delta y, \Delta z)$ and a rotation θ about the z-axis. Let the coordinates of corresponding points be given by (x'_i, y'_i, z'_i), $i=1,2,3$. The two sets of coordinates of the points are related by the relationships:

$$\begin{aligned} x'_i &= x_i\cos\theta - y_i\sin\theta + \Delta x \\ y'_i &= x_i\sin\theta + y_i\cos\theta + \Delta y \\ z'_i &= z_i + \Delta z \end{aligned} \tag{7}$$

for $i=1,2,3$. The perspective transformation imposes the constraint that

$$\begin{aligned} \phi_i &= \frac{x_i}{y_i} = \frac{X_i}{F} \\ \eta_i &= \frac{z_i}{y_i} = \frac{Z_i}{F} \end{aligned} \tag{8}$$

where (X_i, Z_i) are the image plane coordinates of the point (x_i, y_i, z_i) and F is the focal length of the perspective projection.

In the derivation of the above relationships, it is assumed that the optical axis is aligned with the y-axis, and the rotation of the object takes place about the z-axis. The geometry of the above relationships is illustrated in Figure 3.

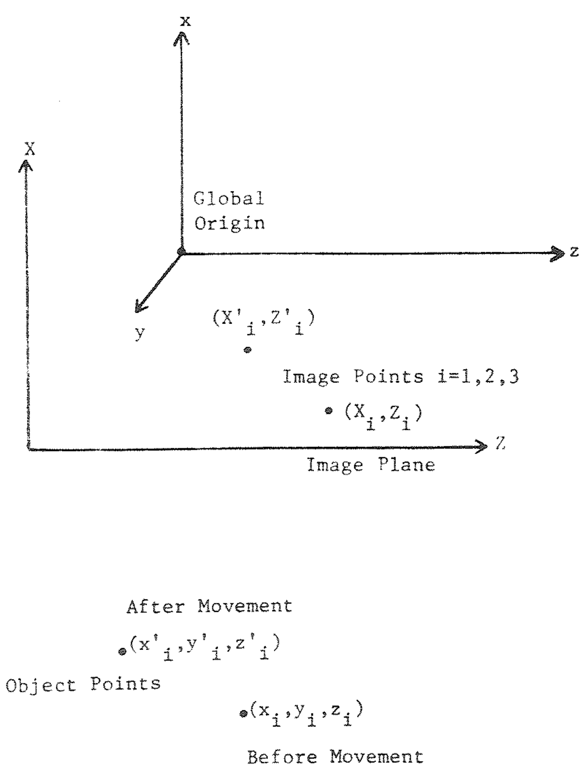

Fig. 3. Geometry for Ullman's polar equation.

On combining the perspective and motion constraints, one obtains

$$\begin{aligned} \phi'_i &= \frac{x'_i}{y'_i} = \frac{x_i\cos\theta - y_i\sin\theta + \Delta x}{x_i\sin\theta + y_i\cos\theta + \Delta y} \\ \eta'_i &= \frac{z'_i}{y'_i} = \frac{z_i + \Delta z}{x_i\sin\theta + y_i\cos\theta + \Delta y} \end{aligned} \tag{9}$$

and on substituting

$$x_i = y_i\phi_i \qquad z_i = y_i\eta_i, \tag{10}$$

the above equations reduce to:

$$\begin{aligned} \phi'_i &= \frac{y_i\phi_i\cos\theta - y_i\sin\theta + \Delta x}{y_i\phi_i\sin\theta + y_i\cos\theta + \Delta y} \\ \eta'_i &= \frac{y_i\eta_i + \Delta z}{y_i\phi_i\sin\theta + y_i\cos\theta + \Delta y} \end{aligned} \tag{11}$$

The following observations may be made about these equations:

1. There are six equations in seven unknowns Δx, Δy, Δz, θ, y_1, y_2, y_3.
2. Ullman provides a procedure for reducing the six equations to a single equation in θ.
3. One variable is used to provide the scale.
4. The derived equation, called the polar equation, is of the form
   $A\sin^2\theta + B\cos^2\theta + C\cos\theta\sin\theta + D\sin\theta + E\cos\theta = 0.$
5. In general the polar equation has four roots and consequently there is ambiguity in the choice of θ.

Ullman suggests a possible strategy for settling on the unique solution by introducing redundancy. In particular, by considering four points and taking the common solution corresponding to various triplets, the ambiguity may be resolved. The results of experimental examples show that the correct answer can usually be found from as little as two views of 4 points.

Solution of the polar equation and the correct choice of the root leads to the computation of (x_i, y_i, z_i), (x'_i, y'_i, z'_i), $i=1,2,3$ and $\Delta x, \Delta y, \Delta z$ and θ in terms of the scaling parameter. The assumptions are centered around the rigidity of the object and the motion constraint.

3.2 A Generalization of the Stereo Problem

Roach and Aggarwal consider the problem as illustrated in Figure 4. The object is assumed to be stationary whereas the camera undergoes an unknown motion. In order to reconstruct the motion and the three-dimensional coordinates of object points from two-dimensional images, two views of five points are needed as shown in the following analysis.

The camera position model considered by Roach and Aggarwal is slightly more general than that of Duda and Hart [3]. In addition to the angles θ and ϕ for pan and tilt, they consider the angle κ for the rotation of the image plane coordinates axes relative to the global coordinate axes. This leads to three position coordinates (x_0, y_0, z_0) and three angle coordinates (θ, ϕ, κ) for specifying the camera position. The equations relating the image plane coordinates (X,Y) with the spatial coordinates (x,y,z), focal length F, and the camera parameters (x_0, y_0, z_0), (θ, ϕ, κ) are given in [5].

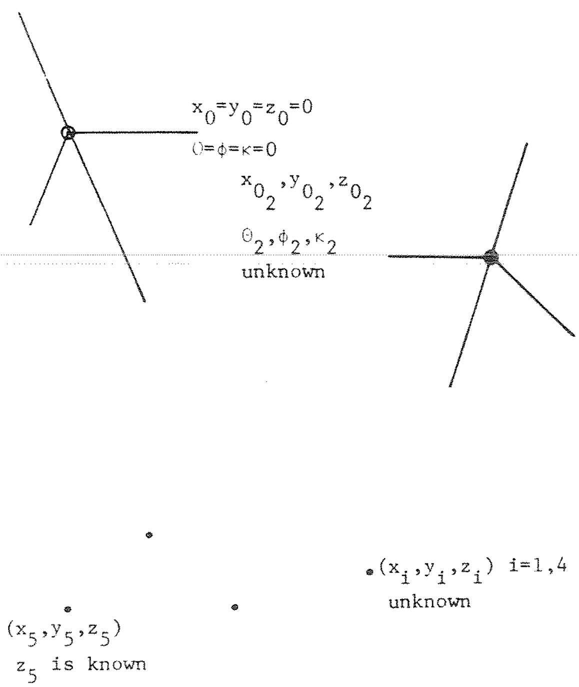

Fig. 4. Roach's camera configuration and point coordinates.

In Figure 4, there are twenty-seven unknowns: fifteen for the three-dimensional coordinates of five points and twelve for the two camera positions (for each position, three for the lens center and three for the optical axis). However, from two views of five points, there are only twenty equations (two for each point in each view). If the six parameters for the first camera position are known (it is convenient to assume them to be zero), and z-coordinates of one of the points is chosen as the scaling factor, the number of unknowns and the number of equations is exactly the same. It is convenient to compute the x and y coordinates of the point, whose z-coordinate is selected as the scaling factor, from the equations $X=Fx/z$ and $Y=Fy/z$. Therefore, there are eighteen equations with eighteen unknown parameters. In general, these eighteen equations present a formidable task for solution. The difficulty of obtaining the solution is considerably aggravated by the presence of noise.

The system of nonlinear projection equations explained above can be solved by using a modified finite difference Levenberg-Marquardt algorithm due to Brown [13-15] without strict descent that minimizes the least-squared error of the 18 equations. The method employed is iterative and requires an initial guess for each unknown parameter.

This work is somewhat like the camera calibration systems of Sobel [16] and Yakimovsky and Cunningham [17]. In their work multiple images of points together with a central projection model and numerical methods are used to determine camera parameters such as focal length, position, and orientation. These studies, however, have considerably more information about the three-dimensional positions of points than we are assuming. Thus, the problems being solved and the information given for the calibration systems are different from the work described in this section.

Implicit in this work are two very important assumptions: that the objects being observed are rigid and that the images of the object are noise free and thus completely accurate. To test the effect of the second assumption on the numerical method described above, from one to four pixels were randomly added to or subtracted from the exact photocoordinate data for a moving object. This perturbation of the data causes extreme instability in the numerical solutions. However, one of the main reasons for using a least-squared error technique to solve a problem is to make adjustments to observations that contain error (noise). Adjustment is only possible, however, when there are more equations than unknowns. Two views of five points are therefore inadequate for noisy data since there are the same number of equations as unknowns. Two views of six points or three views of four points produce 22 equations in 21 unknowns using the same problem model discussed above. Examination of experimental runs using overdetermined systems of equations shows that minimal overdetermination is not very accurate. It is only with considerable overdetermination (two views of 12 or even 15 points; three views of seven or eight points) that the results become accurate.

3.3 A Generalization of the Polar Equation

Nagel uses compact vector notation to formulate the problem of structure and motion. The camera is assumed to be stationary, whereas the object is moving. There are two coordinate systems. One is attached to the camera and the second one to the object. In the camera coordinate system, each object point is expressed as $C_{mi}\vec{c}_{mi}$ where the subscripts i and m refer to the view and point respectively and the vector $\vec{c}_{mi}$ is a unit vector. The same point in the object coordinate system is denoted by $\vec{A}_m$. The object undergoes a translation $\vec{T}$ followed by a rotation R. The relationships for the two views (as shown in Figure 5) are given by

$$C_{m1}\vec{c}_{m1} = \vec{A}_m \qquad (12)$$

$$C_{m2}\vec{c}_{m2} = (\vec{A}_m + \vec{T})R \qquad (13)$$

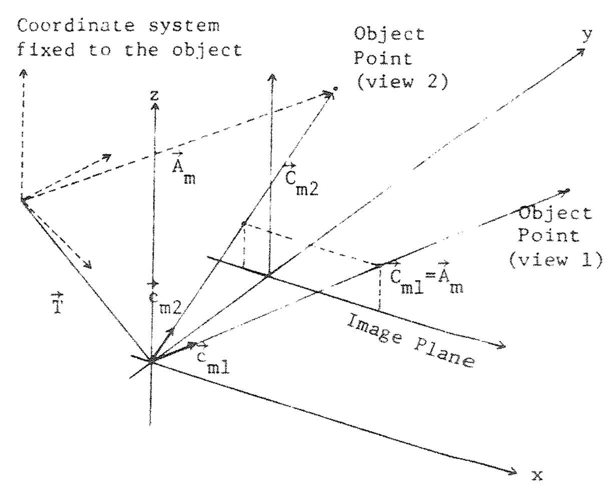

Fig. 5. Nagel's geometry of the two coordinate systems.

In these equations $\vec{c}_{m1}$ and $\vec{c}_{m2}$ are known from the image plane coordinates and lens center location; whereas C_{mi}, $\vec{A}_m$, $\vec{T}$ and R are unknown. It may be noted that in this geometry the lens center is at the origin.

The matrix R can be expressed in terms of the direction cosines of the axis of rotation and the rotation angle. Nagel outlines a procedure by which various unknown parameters may be computed in terms of the scale parameter C_{11}. Nagel makes extensive use of vector algebra in deriving the above results and uses five points in two views to derive the motion and structure parameters. Also, for the special case similar to Ullman, the equation reduces to the polar equation derived earlier. Thus, Nagel's results substantiate earlier work of Ullman and Roach and Aggarwal. Again, the assumption of rigidity of the object and the existence of correspondence between the image points in two views are used to derive the above results.

3.4 The Planar Patch

Tsai and Huang consider an approach significantly different from the above approaches. Instead of considering the motion of individual points, they consider the motion of a planar patch characterized by eight points. The configuration of the coordinate system, camera and object points are shown in Figure 6. The perspective constraints are expressed as:

$$\text{View 1:}\quad X = F\frac{x}{z},\qquad Y = F\frac{y}{z} \tag{14}$$

$$\text{View 2:}\quad X' = F\frac{x'}{z'},\qquad Y' = F\frac{y'}{z'} \tag{15}$$

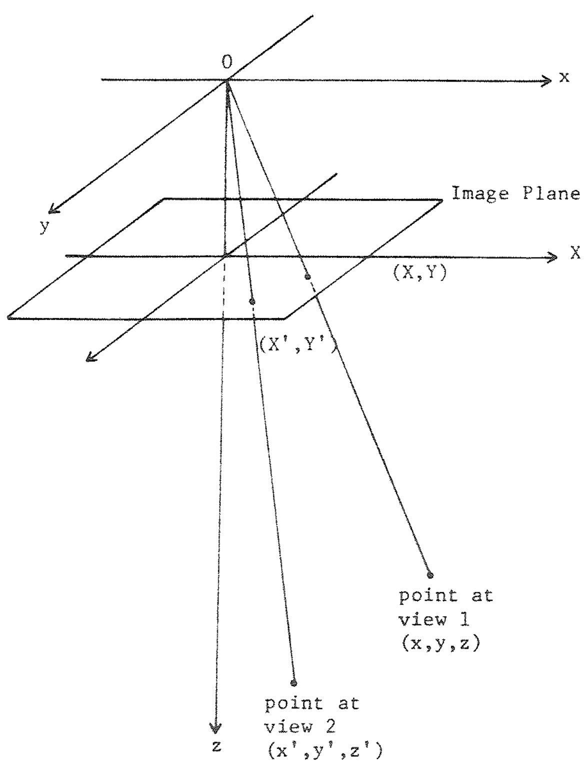

Fig. 6. Tsai and Huang geometry of the image plane.

The motion of a point is expressed as

$$\begin{vmatrix} x' \\ y' \\ z' \end{vmatrix} = R \begin{vmatrix} x \\ y \\ z \end{vmatrix} + \vec{T}$$

where R is the rotation matrix and $\vec{T}$ is the translation vector. Tsai and Huang assume that there is a rigid planar patch in the object space characterized by eight points and given by the equation

$$ax + by + cz = 1 \tag{16}$$

Further, they express the change of position of points in the image plane as the transformation

$$X' = \frac{a_1X+a_2Y+a_3}{a_7X+a_8Y+1}\qquad Y' = \frac{a_4X+a_5Y+a_6}{a_7X+a_8Y+1} \tag{17}$$

where $a_i, i=1,\ldots,8$ are called the pure parameters. Using the motion, rigidity and perspective constraints, the pure parameters may be expressed in terms of the planar patch and motion parameters. After a tedious algebraic manipulation a sixth order polynomial may be formulated such that its coefficients are expressed in terms of pure parameters and its solution yields (together with additional manipulation) the motion and structure parameters in terms of scale. The sixth order polynomial appears to have only two real roots, however, no analytical demonstration of this fact is possible at this time.

3.5 A Coherent View

The commonality in the underlying assumptions and the analyses of the four papers reviewed above may be summarized as follows:

1. Rigidity of the object
2. Perspective Transformation
3. Combining perspective and motion constraints
4. Solution being expressed in terms of a scale factor
5. Need to establish correspondence of points between views.

The solution of the equations, in general, is complex but manageable. However, it is the need to establish correspondence of points between views that has led us to the work described in the next section.

4.0 OCCLUDING CONTOURS IN DYNAMIC SCENES

In this section we will describe the results of work [8] done in the pursuit of two major goals. The first goal is the development of a dynamic scene analysis system that does not depend completely on feature point measurements. The second goal is the development of a scheme for representing three-dimensional objects that is descriptive of surface detail, yet remains functional in the context of structure from motion in dynamic scenes.

To lessen the dependency on feature point detection, occluding contours with viewpoint specifications are used. The term "occluding contour" means the boundary in the image plane of the silhouette generated by an orthogonal projection. Silhouettes can most often be formed by a simple thresholding of the intensity values. A connected component analysis [18, pp. 336-347] of the resulting binary image yields the boundary of the object silhouette. An ordered list of the image plane coordinates of the resulting boundary constitutes the initial representation of the occluding contour. Throughout the analysis of the dynamic image, however, another representation, referred to as the rasterized area [19], of the contour will also be used. For its use here, the most significant attribute of this representation is that given an area so represented and an arbitrary segment on a line parallel to the "raster direction" it is a simple process to determine what portions of the given segment intersect the area, i.e., to clip that segment to the represented area.

The three-dimensional structure to be derived from the sequence of occluding contours is a bounding volume approximation to the actual object. For this reason the representation incorporated in this system is based on volume specification through a "volume segment" data structure. The volume segment representation is a generalization to three-dimensions of the rasterized area description. For the rasterized area, each of the segments denoted a rectangular area. The generalization to three dimensions is to have each segment represent a volume, i.e., a rectilinear parallelepiped with faces parallel to the coordinate axes. In addition to grouping collinear segments into lists, the set of segment lists is partitioned so that the subsets contain lists having coplanar segments. The primary dimension of the parallelepiped specified by a segment is the length of the segment. The second dimension is given by the inter-segment spacing within the plane of the segment, while the third dimension is the inter-plane distance. The latter two dimensions are specified to be uniform throughout the volume segment representation.

The primary advantage of this structure in general applications is that the process of determining whether an arbitrary point is within the surface boundary consists of a simple search of three ordered lists: select a "plane" by z-coordinate; select a "line" by x-coordinate; and finally, check for inclusion of the y-coordinate in a segment.

This volume segment representation is created from a dynamic image by two processes. The first process combines information from frames 1 and 2 of the dynamic image to form an initial volume segment representation. The second process then accepts each succeeding frame in order to refine the approximation represented by the volume segment structure. Thus these processes analyze the occluding contours with their view orientations to initially construct and to continually refine the volume segment representation of the object generating the contours. Algorithm summaries of the two processes are given in more detail in [20].

Two examples illustrate the results possible by this method.

Example 1. Four silhouettes of a box with knobs are shown in Figure 7. On combining occluding contours, labelled a and c; a,c and b; and a,c,b and d, one obtains successively the volumes shown in Figures 8,9 and 10. The continuing refinement of the process is fairly apparent.

Example 2. Figure 11a,b,c give the occluding contours for a rectangular parallelpiped with a hole. Figure 12 and 13 give the volume constructed from a and b; and a,b and c respectively.

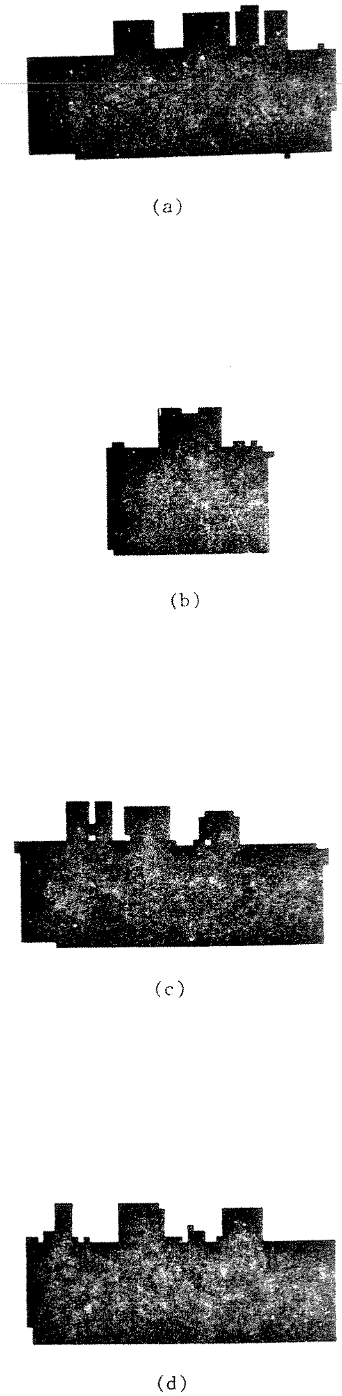

Fig.7. Silhouettes for a box from four different viewpoints.

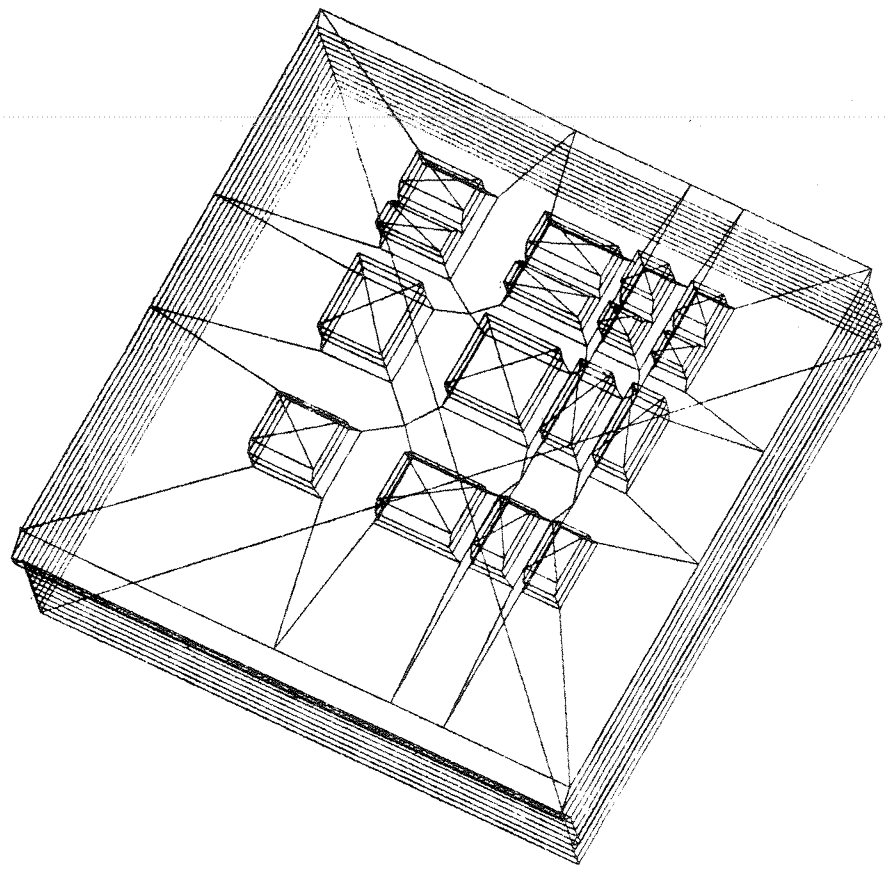

Fig. 8. Box surface description based upon 7(a) and (c).

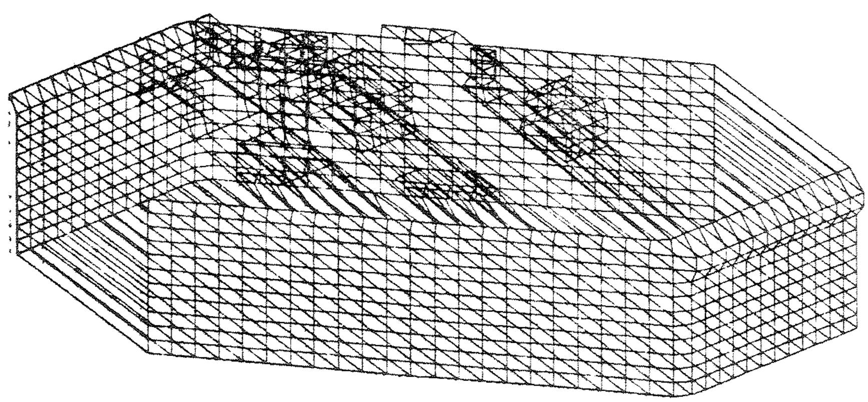

Fig. 9. Box surface description based upon 7(a), (c) and (b).

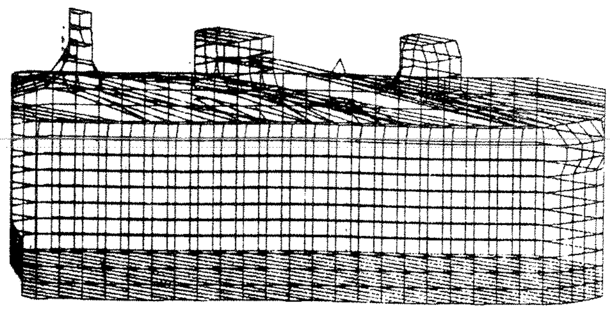

Fig. 10. Box surface description based upon 7(a), (c), (b) and (d).

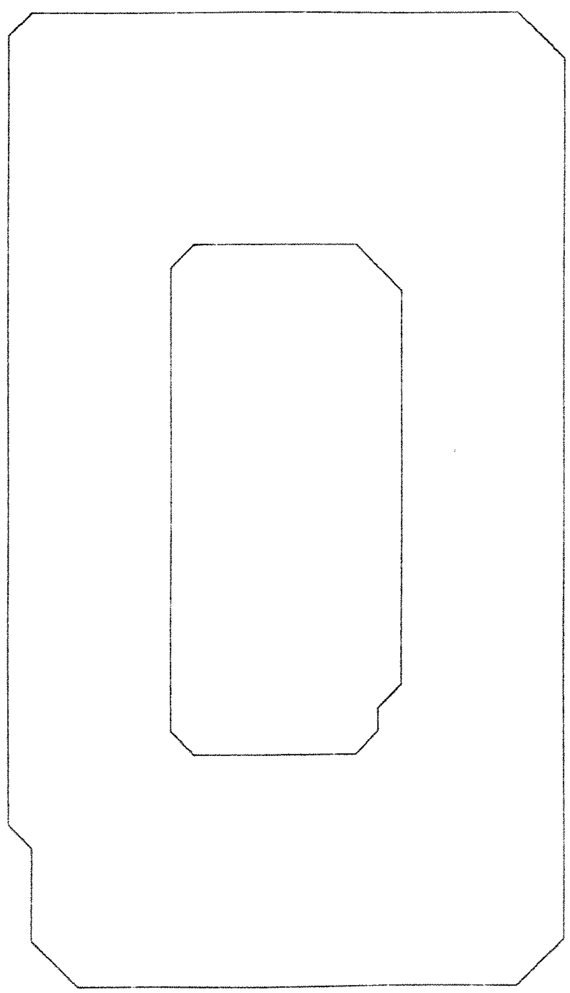

Fig. 11 (a). Occluding contours for rectangular parallelpiped with a hole.

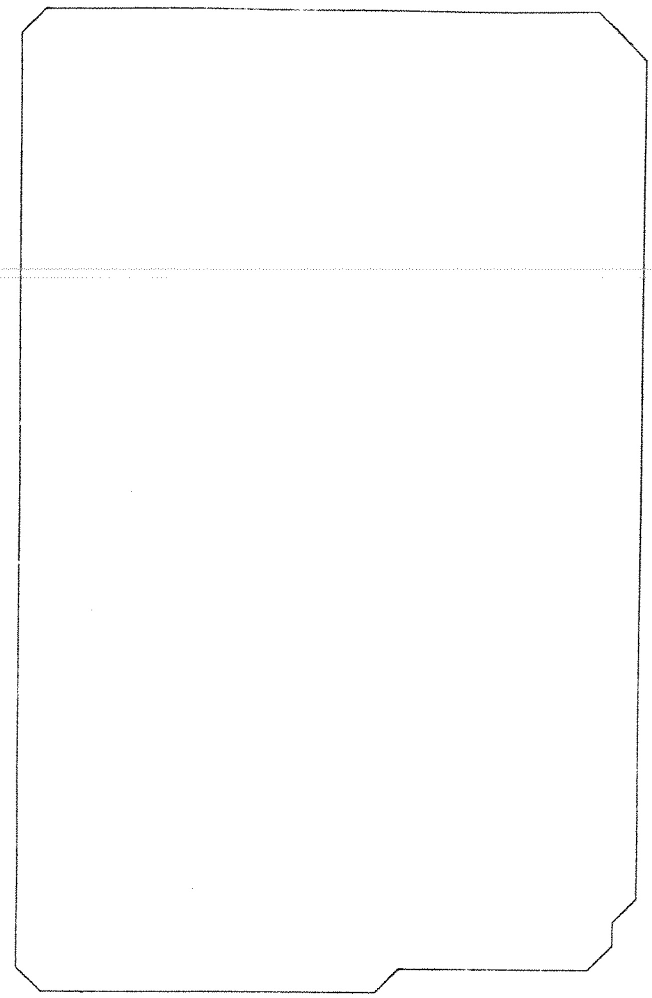

Fig. 11(b). Occluding contours for rectangular parallelpiped with a hole.

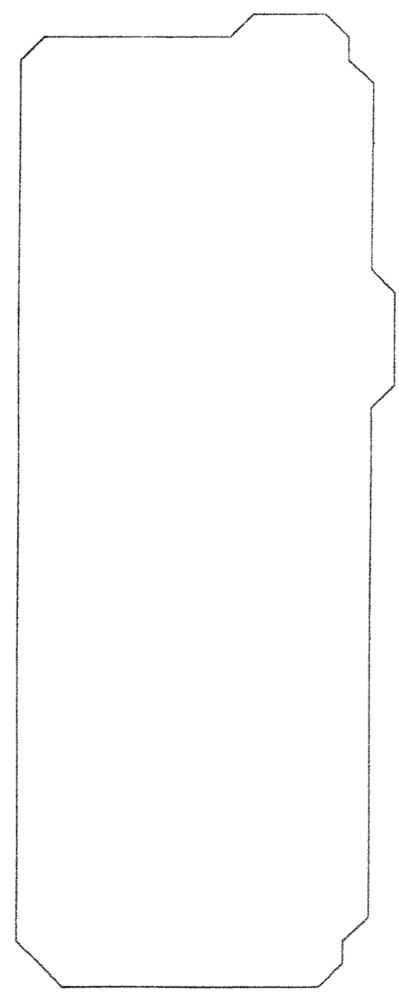

Fig. 11(c). Occluding contours for rectangular parallelpiped with a hole.

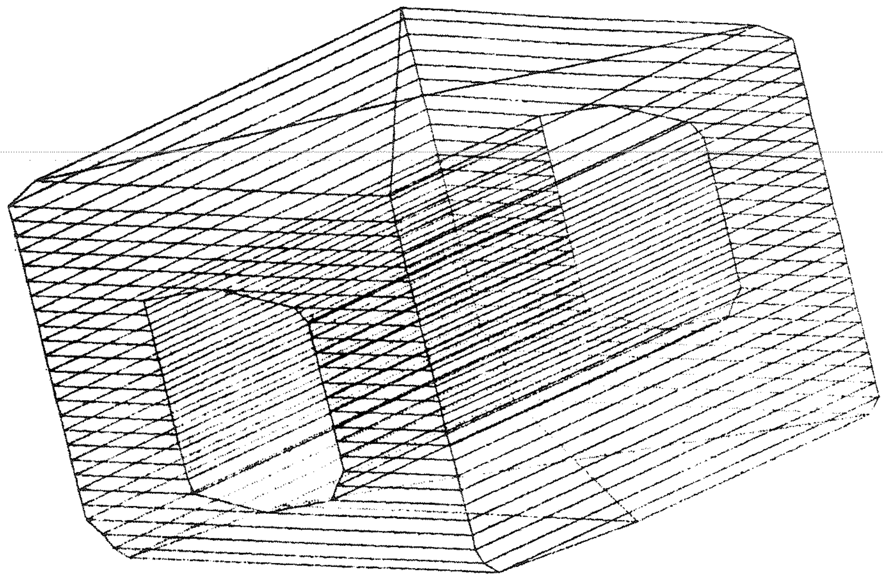

Fig. 12. Surface description based on 11(a) and (b).

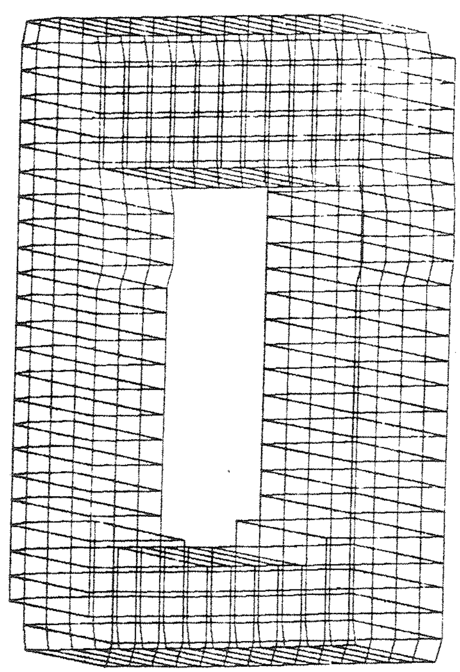

Fig. 13. Surface description based on 11 (a), (b) and (c).

5.0 FUTURE DIRECTIONS FOR RESEARCH

The review of previous sections has focused on the computation of three-dimensional structure and the motion vector of objects moving in space from two-dimensional images. The need for efficient methods for inversion of perspective equations and for storage of three-dimensional description of objects becomes fairly obvious. The problems related to the three-dimensional description of objects have been around for a long time and they have received the attention of some more creative researchers. However, there is still a need for additional research in this area. The properties of an "ideal" three-dimensional description of objects are fairly easy to enumerate but this ideal is far from being achieved. In particular, an ideal description should have the following properties:

i. The description should be compact so that it is amenable to easy storage and fast transmission.

ii. The description should be easily transformable to viewer-centered description.

iii. The description should be readily convertible to other object-centered three-dimensional descriptions.

iv. The description should be amenable for partial description of objects and easy to update when additional information is available.

v. The description should be able to accomodate holes and concavities.

vi. The description should be extendable to include deformation of objects.

The intensive computational needs for inversion of perspective equations in a noisy environment are rather severe. If one adds the constraints for real time processng of video images, the computational needs are indeed astronomical. Not only must one provide for a super computer, but one must come up with rather creative and innovative solutions to the numerical drudgery. In particular, the real time processing of color video images of 512x512 with 8 bits intensity for each color requires the processing of approximately 2^{26} bits/sec. Parallel processing of data and its early reduction are important, necessary ingredients.

References

1. J. K. Aggarwal, L. S. Davis, W. N. Martin and J. W. Roach, Survey: Representation Methods for Three-Dimensional Objects, in L. N. Kanal and A. Rosenfeld, (eds), Progress in Pattern Recognition, Vol. 1, (North-Holland Publishing Company, New York, 1981) 377-391.

2. B. Gil, A. Mitiche, and J. K. Aggarwal, Experiments in Combining Intensity and Range Edge Maps, to appear in Computer Graphics and Image Processing.

3. R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis, (John Wiley and Sons, New York, 1973).

4. S. Ullman, The Interpretation of Visual Motion, (MIT Press, Cambridge, MA., 1979).

5. J. W. Roach and J. K. Aggarwal, Determining the Movement of Objects from a Sequence of Images, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, No. 6, (Nov. 1980) 554-562.

6. H.-H. Nagel, Representation of Moving Rigid Objects Based on Visual Observations, Computer, Vol. 14, No. 8, (Aug. 1981) 29-39.

7. R. Y. Tsai and T. S. Huang, Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch, IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-29, No. 6, (Dec. 1981) 1147-1152.

8. W. N. Martin and J. K. Aggarwal, Volumetric Descriptions of Objects from Multiple Views, to appear in IEEE Transactions on Pattern Analysis and Machine Intelligence.

9. W. N. Martin and J. K. Aggarwal, Dynamic Scene Analysis: A Survey, Computer Graphics and Image Processing, Vol. 7, No. 3, (June 1978) 356-374.

10. J. K. Aggarwal and W. N. Martin, Dynamic Scene Analysis, Technical Report 82-3, Laboratory for Image and Signal Analysis, The University of Texas Austin, Texas, (Sept. 1982).

11. J. K. Aggarwal, L. S. Davis and W. N. Martin, Correspondence Processes in Dynamic Scene Analysis, Special Issue on Image Processing, Proceedings IEEE, Vol. 69, No. 5, (May 1981) 499-657.

12. J. A. Webb and J. K. Aggarwal, Shape and Correspondence, to appear in Computer Graphics and Image Processing.

13. K. M. Brown and J. E. Dennis, Derivative Free Analogues of the Levenberg-Marquardt and Guass Algorithms for Nonlinear Least Square, Numer. Math., Vol. 18, (1971) 289-292.

14. K. Levenberg, A Method for the Solution of Certain Nonlinear Problems in Least Squares, Quart. Appl. Math., Vol. 2, (1944) 164-168.

15. D. W. Marquardt, An Algorithm for Least Squares Estimation of Nonlinear Parameters, J. SIAM, Vol. 11, No. 2, (1963).

16. I. Sobel, On Calibrating Computer Controlled Cameras for Perceiving 3-D Scenes, Artificial Intelligence Vol. 5, No. 2, (1974) 185-198.

17. Y. Yakimovsky and R. Cunningham, A System for Extracting Three-Dimensional Measurements from a Stereo Pair of TV Cameras, Computer Graphics and Image Processing, Vol. 7, No. 2, (Apr. 1978) 195-210.

18. A. Rosenfeld and A. C. Kak, Digital Picture Processing, (Academic Press, New York, 1976).

19. W. M. Newman and R. F. Sproull, Principles of Interactive Computer Graphics, second edition, (McGraw-Hill, New York, 1979).

20. W. N. Martin and J. K. Aggarwal, Analyzing Dynamic Scenes, Technical Report 81-5, Laboratory for Image and Signal Analysis, The University of Texas , Austin, Texas, (Dec. 1981).

Acknowledgments: It is a pleasure to acknowledge the help of Messrs. C. H. Chien, B. Gil, Y. C. Kim and Dr. W. N. Martin during the preparation of this paper.

This research was supported by the Air Force Office of Scientific Research under Contract F49620-83-K-0013.