Model-based Recognition of Human Posture Using Single Synthetic Images.

C. I. Attwood[§], G. D. Sullivan & K. D. Baker.

Intelligent Systems Group, Dept. of Computer Science,
University of Reading, RG6 2AX, UK.
Charlie.Attwood@reading.ac.uk

Model-based vision has been predominantly concerned with the recognition of single component, rigid objects. This paper describes work attempting to recover the 3D structure of a multi-component, highly articulated object - the human body. A major goal of this research has been to go beyond the level of basic object recognition, and attempt to reach a semantic level of description regarding the object's behaviour. Previous research on the recognition of human figures has assumed that the behaviour of the figure in the image is known a priori, *for example, "walking", or has made use of motion information derived from image sequences. This research shows that accurate 3D structure can be recovered without such knowledge, and that descriptions of a human figure's behaviour can be obtained in terms of static posture descriptions such as; "sitting", "kneeling", "standing".*

The problem of human figure recognition has received attention from the vision community during the last decade, [1,2,7,8,9,10,12,13,15]. The human body represents a difficult object to recover since it has multiple components and each linkage point possesses a high degree of motility. Previous research has followed two themes:

- High performance at lower level vision tasks e.g. segmentation/labelling is usually accompanied by little or no attempt at solving higher level problems, such as recovery of 3D structure.
- Conversely, success at recovery of higher level information is usually predicated upon simplifying assumptions at lower levels.

The present research falls within the latter category; it uses schematic synthetic image data, and seeks to recover 3D human posture descriptions.

Previous work on human figure recognition has required **motion** information, *prior* to recovery of the correct physical structure, either in the form of knowledge of the behaviour of the figure (e.g. walking), or as general motion information derived from image sequences. Such knowledge greatly reduces the difficulty of the recovery problem since it allows a number of powerful constraints to be imposed. Motion information has been used in previous research, in at least one of the following ways:

- structure from motion, [12,15].
- using *a priori* knowledge of the complex motion of the figure and/or its sub-parts to make predictions about subsequent frames, [1,2,8,9,10].
- using rules of motion in single images, based on *a priori* knowledge of the behaviour a person is engaged in, i.e. walking, [7].

We report here a method which achieves detailed structure recovery *without* using prior knowledge of a figure's behaviour or motion information. Knowledge of human anatomy and posture can provide sufficient information to allow recovery of accurate 3D structure from static images of human figures. The structure information obtained is then used as the basis from which *posture recognition* is achieved.

THE BODY MODEL

The basic model consists of a collection of 16 cylinders (body segments) and 14 joints. The connectivity of body parts is represented as a tree structure. Each segment and joint possesses its own rigidly embedded right-handed local coordinate system. All joints have three axes of rotation, except the hinge joints (elbows and knees), which are single axis. In a complete model one segment is specified as a *root segment* (the root of the tree). This segment is used as the starting coordinate system from which the transformations for the rest of the model's joints and segments are concatenated. The *root segment* therefore forms the origin of a model-centred coordinate system and, as such, can be used to position the model in 3-space. The model possesses a flexibility comparable with a fit adult human - anatomical limits for each axis of each of the body model's joints were obtained from reference texts which specify the possible degrees of limb movement

§ Research funded under contract from IBM (UK), Scientific Centre, Winchester.

AVC 1989 doi:10.5244/C.3.5

for use in reconstructive surgery, [4-6].

The body model serves two major and quite distinct functions:

- **Graphical:** As an instance *generator* to simulate image data.
- **Interpretation:** As a parameterised representation, which supports model-based interpretation of image data.

The model makes use of several types of knowledge ranging from physical to behavioral, including; mass, length, volume, connectivity, anatomical limits on joint movement, constraints due to postures, collision detection and, in the case of a standing posture, balance testing.

SYNTHETIC IMAGE DATA

Previous research, which has concentrated on low level processing of real images, has shown some success in finding and labelling body parts, for example, by detecting ribbons (c.f. [3]) in the image, [1,8,9]. It is reasonable to assume that the positions of joints in an image could be found using, for example, the intersection of 2D ribbon axes. The present research makes the assumption that the position in the image of the major joints can be identified, and seeks to recover the full 3D position of the figure. This joint position data represents the sole source of information (with the exception of knowledge of the position of the root segment, see below) used by the structure recovery process.

DEFINING POSTURES

Postures are defined by specifying permissible joint angle ranges. Within a single posture definition, the angle range over which each axis of a joint may vary is expressed in the form of a probability density function (pdf). Each pdf describes the relationship between the angle of an axis of a joint and its likelihood of occurrence relative to the range of permissible angles. The pdfs are used in two different forms in the graphical and interpretative phases respectively:

- To generate an angle with a probability as defined by the pdf.
- To return a likelihood value for a given angle.

Figure 1. illustrates the angle generator usage for the x-axis of the right hip in a standing posture - angles near 0° are most likely, but angles up to ±10° are possible. The pdfs are defined as histogram functions of between 3 and 5 steps, these are then converted into lookup tables for each of the two uses described above. Figure 2. shows examples of the three major posture types as generated using the pdfs.

All generated postures are tested for illegal collision between body segments, before being used to provide synthetic image data. In addition, standing model instances have their feet fixed to the ground plane, and are correctly balanced - that is, the projection of the model's centre of gravity lies within the convex hull of the foot area.

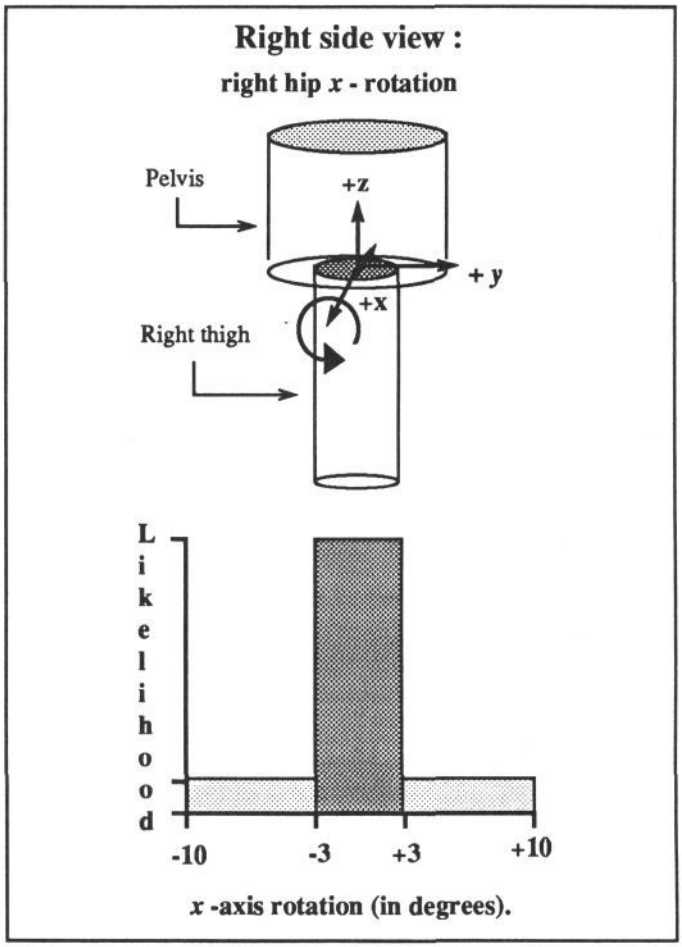

Figure 1. The pdf definition for the x-axis of the right hip in the posture of standing.

STRUCTURE RECOVERY

The position of joints in the image is the main input to the system, and no use is made of other cues such as stereo, motion, shape from shading, *etc*. The recovery of the precise 3D structure of a human figure from our 2D image data is grossly underdetermined. This problem (perspective inversion) can only be solved by invoking extra knowledge sources. One source of this extra knowledge is the object model used. Information about limb length, the planar relationship of certain joints, and anatomical limits on joint movement, may be used. Other sources included; non-collision between body parts, the laws of perspective, and balance (in the case of a standing figure).

The input to the recovery algorithm was the (correctly labelled) 2D image coordinates of the major joints, and the

transformation linking the camera coordinate system with the coordinate system of one body segment (the *root segment*). Knowledge of the (2D) end-points of a body segment of known (3D) length partially constrains the 3D position of the segment. A method devised by Webb[14], and adapted by Lee & Chen[7] was further modified in order to exploit this fact. The position in depth of a single joint embedded in the root segment was obtained and the positions of connected segments were expanded from this starting point. There are two possible solutions for each body segment, corresponding to a foreshortening due either to a forward or backward tilt of the segment. A complete expansion of the problem space therefore forms a binary tree of depth equal to the number of body segments.

The body model possesses 14 joints, giving 8,192 possible body configurations using this information alone (2^{n-1}, where n = no. of joints). This figure corresponds to the search task when only the *positions* of joints in 3-space are required. The true search space for the instantiated model involves 262,144 (2^{18}) possible body configurations, since the position of the tips of body extremities (head, hands & feet) must also be found, to allow recovery of the *angles* of the extremal joints.

Constraints derived from the knowledge sources described above, were used to prune the binary tree. Anatomical limits provide the greatest pruning power. This requires the recovery of precise joint angles at each level of the binary tree's expansion. At each node expansion of the binary tree the joint transformation matrix which links connected segment pairs was recovered, using geometric reasoning methods, partly in the form of a modification of the "*rotation about an arbitrary vector*" technique [11]. Each matrix was then decomposed to recover the angles between limb vectors, expressed as axis rotations in the joint's local coordinate system. All angles were recovered to an accuracy of better than 1 degree. A new method for uniquely decomposing the transformation matrix was developed. This method is applicable in cases where one axis of rotation lies within $\pm\pi/2$ and the other two axes within $\pm\pi$. The $\pm\pi/2$ limit, corresponds to the natural anatomical limit for limb rotation along the major axis, allowing the method to be applied to the three-axis (and single-axis) joint decomposition problem (see appendix A).

The end result of a fully expanded and pruned binary tree is one or more leaf nodes each representing a separate interpretation of the 3D structure of the figure in an image.

POSTURE RECOGNITION

Three major postures have been defined (standing, sitting and kneeling) together with two sub-postures - "reach-left" and "reach-right". Posture recognition therefore includes compound descriptions, such as, "sitting-reach-left".

Posture recognition is achieved using the pdf lookup table format which returns a likelihood for a given joint axis' rotation value, as described earlier. The algorithm works as follows:

For each anatomically correct ***Body Interpretation***
  For each ***Posture*** in the Database
    For each ***Joint*** in a ***Posture***
      Obtain the angle for each axis of the current ***Joint.***
      Obtain ***Joint's*** Likelihood value for the current ***Posture.***
      If the likelihood value of a ***Joint*** axis = zero **then**
        quit current ***Posture.***
      else
        Accumulate the likelihood value for current ***Joint.***
      end
    Next ***Joint***
    Store ***Posture*** as recognised (with average likelihood value).
  Next ***Posture***
Next ***Body Interpretation***

Interpretations for which no posture description is found are rejected. The output of the algorithm is a list of surviving body interpretations each with a posture description and associated average likelihood value.

Collision Detection and Balance Constraints

All surviving interpretations are also tested to check that the structure interpretation does not involve illegal collision between body segments. This test is performed after posture recognition so that advantage may be taken of the general relationship between body parts in a given posture. For example, hands and feet only need be tested for inter-collision, when a kneeling posture has been recognised. Any body interpretation in which a collision is detected can be rejected. When a standing posture is found, any surviving interpretations are also tested for balance and interpretations failing to balance are also rejected.

EXPERIMENTAL RESULTS

Table 1. shows how anatomical pruning controls the binary tree's growth at each joint expansion node. The results for 36 synthetic images are presented in tables 2 - 4. The correct posture description was obtained in every case. The number of interpretations remaining after all forms of pruning had been applied ranged between 2 and 48 per image (mean = 7.03), and in all cases one interpretation held the correct structure. Many of these multiple solutions actually arise from very small ambiguities in depth, for example, a hand tipped towards or away from the camera. These ambiguities can combine such that, for example, 3 trivial depth choices can give rise to 8 interpretations. There is an inherent ambiguity in the case where an extremal segment has two non-colliding, anatomically legal interpretations - this potential indeterminacy is unavoidable unless additional feature

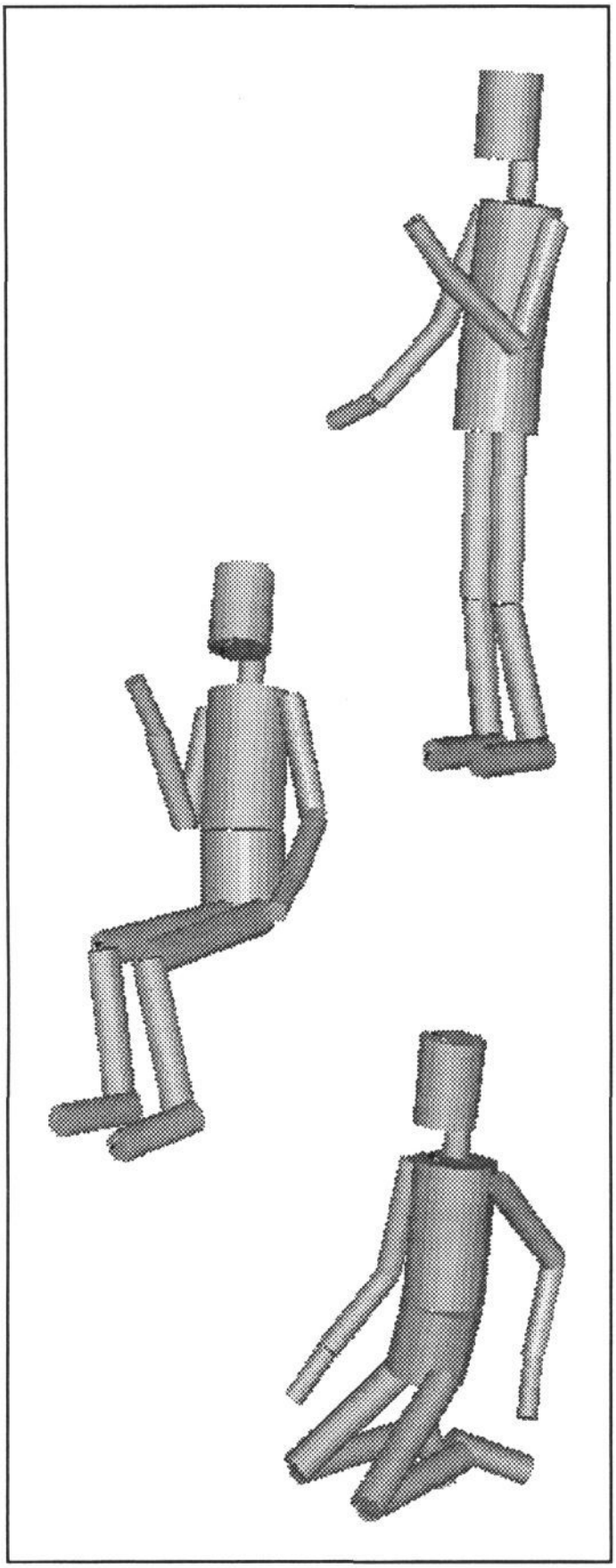

Figure 2. Examples of the pdf defined postures: Standing (image no. 3.), Sitting (image no. 15) and kneeling (image no. 34).

Level number	Number of interpretation tree paths.	
1	Expand left hip	2
2	Expand left knee	4
	Prune left hip	2
	Prune left knee	2
3	Expand left ankle	4
	Prune left ankle	2
4	Expand right hip	4
5	Expand right knee	8
	Prune right hip	4
	Prune right knee	2
6	Expand right ankle	4
	Prune right ankle	2
7	Expand lumbar	4
8	Expand lower cervical (to L-shoulder)	8
	Expand lower cervical (to R-shoulder)	16
	Prune lumbar and L&R shoulders *	2
	Expand lower cervical (to head tip)	4
	Prune lower cervical	2
9	Expand left shoulder	4
10	Expand left elbow	8
	Prune left shoulder	4
	Prune left elbow	2
11	Expand left wrist	4
	Prune left wrist	2
12	Expand right shoulder	4
13	Expand right elbow	8
	Prune right shoulder	6
	Prune right elbow	2
14	Expand right wrist	4
	Prune right wrist	4

* a constraint based on distance between the shoulders is used to prune both the shoulder positions and the lower cervical position.

Table 1. Growth and pruning of the binary tree (image no. 15).

points are identified on the segment. Since none of the ambiguities remaining prevent the correct posture description being found, those which arise at extremal joints or have average depth errors less than the equivalent of 5cms are excluded. The final number of surviving interpretations per image ranges between 1 and 3, (mean = 1.333).

EXTENSIONS TO RESEARCH

An investigation of the sensitivity of the structure recovery method to error is currently under way. The aim is to emulate the kind of joint location error that low level segmentation processes might produce. Early results suggest that the technique will be viable using the quality of data likely to be obtainable from real images.

Image Number	1	2	3	4	5	6	7	8	9	10	11	12
After Anatomical Pruning	12	16	48	12	3	5	4	16	6	16	4	32
After Posture Recognition	6	8	48	6	3	5	4	4	3	8	4	8
After Collision Detection	4	4	48	6	3	5	4	4	3	8	4	8
After Trivial Ambiguity Removed	**1**	**1**	**1**	**1**	**2**	**1**	**1**	**2**	**1**	**1**	**1**	**2**

Table 2. Standing - Cells contain the number of surviving whole body interpretations.

Image Number	13	14	15	16	17	18	19	20	21	22	23	24
After Anatomical Pruning	4	8	4	4	24	3	6	24	6	2	32	8
After Posture Recognition	4	8	2	4	24	3	6	12	6	2	16	8
After Collision Detection	4	8	2	2	12	3	6	12	6	2	16	6
After Trivial Ambiguity Removed	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**	**1**	**1**	**1**

Table 3. Sitting - Cells contain the number of surviving whole body interpretations.

Image Number.	25	26	27	28	29	30	31	32	33	34	35	36
After Anatomical Pruning	2	4	2	10	2	48	5	3	4	18	24	12
After Posture Recognition	2	4	2	10	2	12	5	3	4	18	12	6
After Collision Detection	2	2	2	5	2	12	5	3	4	18	12	6
After Trivial ambiguity removed	**1**	**1**	**1**	**2**	**1**	**2**	**1**	**1**	**3**	**2**	**3**	**2**

Table 4. Kneeling - Cells contain the number of surviving whole body interpretations.

CONCLUSION

A method has been described in which naive physics, and knowledge about human anatomy and posture allow both the full 3D structure of a figure to be recovered and a semantic level of description of its behaviour to be obtained. The fact that this is achieved without the use of any form of motion information, should allow the system to cope with the real world case where a human figure is inactive, for example, sitting.

The pdf based definition of posture may have some psychological validity. We have observed that the human visual system often appears to interpret structure on the basis of preconceptions about likely postures. For example, in a silhouette image of a human figure with an extended arm, the human observer "prefers" an interpretation in which the arm is pointing towards the viewer, rather than awkwardly bent backwards. This tendency is reflected, by the probabilities selected for the pdf functions, which are weighted towards "natural" or relaxed stances. Thus, our system would also select the "arm towards the viewer" case, as the best structure interpretation.

APPENDIX A

During the process of structure recovery an homogeneous matrix is derived for each joint, which describes that joint's rotation, i.e. the transformation linking the local coordinate systems of the body segments connected by the joint. To achieve posture recognition it is necessary to recover from this matrix the precise rotation parameter (θ) for each of the joint's axes of rotation.

Firstly we arbitrarily assume that joint axes rotations (R) are concatenated to form the matrix (M), in the fixed order :

$$M = Rz(\theta_z)\ Rx(\theta_x)\ Ry(\theta_y).$$

This expands to :

$$M = \begin{bmatrix} CzCySzSxSy & SzCx & Cz\text{-}SySzSxCy & 0 \\ -SzCyCzSxSy & CzCx & -Sz\text{-}SyCzSxCy & 0 \\ CxSy & -Sx & CxCy & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where :

S = sine, e.g. $Sx = \sin\theta_x$ and

C = cosine, e.g. $Cx = \cos\theta_x$.

If we assume :

$$-\pi/2 < Rz(\theta_z) < \pi/2 \qquad (1.1)$$

$$-\pi \leq Rx(\theta_x) \leq \pi \qquad (1.2)$$

$$-\pi \leq Ry(\theta_y) \leq \pi \qquad (1.3)$$

then the three rotation parameters of M, (θ_x, θ_y, θ_z) can be obtained in the following way:

$$\sin\theta_x = -M(3,2)$$

$$\cos\theta_x = \pm\sqrt{1 - \sin^2 x}$$

Assumption (1.1) defines the cosine of z as positive. It follows that the sign of $\cos\theta_x$ must equal the sign of $CzCx$, which can be obtained from M(2,2). Knowing the correct value for $\cos\theta_x$ the following values can now be obtained:

$$\sin\theta_y = M(3,1) / \cos\theta_x$$

$$\cos\theta_y = M(3,3) / \cos\theta_x$$

$$\sin\theta_z = M(1,2) / \cos\theta_x$$

$$\cos\theta_z = M(2,2) / \cos\theta_x.$$

Knowing the sine and cosine values for each axis of rotation, a trigonometric function utilising quadrant information, can be applied to obtain the correctly signed rotation angles.

REFERENCES

1. **Akita, K.** "Image Sequence Analysis of Real World Human Motion" *Pattern Recognition*, Vol. 17, No.1, (1984) pp. 73-83.
2. **Hogg, D.C.** "Model Based Vision: A Program to See a Walking Person" *Image and Vision Computing*, Vol. No.1, (1983) pp. 1-20.
3. **Binford, T. O.** "Visual Perception by a Computer" *I.E.E.E Conference on systems and controls*, Miami, Florida, December (1971).
4. **Kapandji, I.A.** *The Physiology of the Joints, Vol.1 Upper Limb*, Churchill Livingstone, London (1974a).
5. **Kapandji, I.A.** *The Physiology of the Joints, Vol.2 Lower Limb*, Churchill Livingstone, London (1974b).
6. **Kapandji, I.A.** *The Physiology of the Joints, Vol.3 The Trunk and the Vertebral Column*, Churchill Livingstone, London (1974c).
7. **Lee, H.J. & Chen, Z.** "Determination of 3D Human Body Postures from a Single View" *Computer Vision, Graphics and Image Processing*, Vol. 30, (1985) pp. 148-168.
8. **Leung, M.K. & Yang, Y.H.** "Human Body Motion Segmentation in a Complex Scene" *Pattern Recognition*, Vol. 20, No.1, (1987a) pp. 55-64.
9. **Leung, M.K. & Yang, Y.H.** "A Region Based Approach for Human Body Analysis" *Pattern Recognition*, Vol. 20, No. 1, (1987b) pp. 321-339.
10. **O'Rourke, J. & Badler, N.I.** "Model-Based Image Analysis of Human Motion Using Constraint Propagation" *I.E.E.E Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, No.6, (1980) pp. 522-536. November.
11. **Paul, R.P.** *Robot Manipulators: Mathematics, Programming, and Control.* The M.I.T Press, Cambridge, MA.
12. **Rashid, R.F.** "Towards a System for the Interpretation of Moving Light Displays" *I.E.E.E Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, No.6, (1980) pp. 574-581. November.
13. **Tsukiyama, T. & Shirai, Y.** "Detection of the Movements of Persons from a Sparse Sequence of TV images" *Pattern Recognition*, Vol.18, No's. 3/4, (1985) pp. 207-213.
14. **Webb, J. A.** "Static Analysis of Moving Jointed Objects" *Proc. Amer. Assoc. Artif. Intelligence (AAAI)*, Vol. 1, (1980) pp. 35-37.
15. **Webb, J.A. & Aggarwal, J.K.** "Visually Interpreting the Motion of Objects in Space" *Computer*, Vol. 14, No. 8, (1981) pp. 40-46.
16. **Webb, J.A. & Aggarwal, J.K.** "Structure from Motion of Rigid and Jointed Objects" *Artificial Intelligence*, Vol.19, (1982) pp. 107-130.