Structure from Motion of Rigid and Jointed Objects*

Jon A Webb
J K Aggarwal

Departments of Computer Sciences and Electrical Engineering
The University of Texas at Austin
Austin, Texas 78712

ABSTRACT

A method for structure from motion is presented. The method makes a motion assumption about the objects being viewed. The motion assumption is that all motion consists of translations and rotations about a fixed axis. Parallel projection is also assumed. This makes it possible to interpret the motion of as few as two rigidly connected points. The method works for both rigid and jointed objects. Results of a test of this method on Johansson's data are presented.

1. INTRODUCTION

Human beings have many sources of information on depth that do not depend on stereo vision. These include motion parallax, shape from shading, and textural information. These sources of depth information are particularly interesting for computer applications because they require the use of only one camera. Moreover, there are many situations where stereo algorithms are difficult to use, so that the availability of other sources of depth information is helpful. The present investigation is also motivated for psychological reasons. As computational methods are proposed for solving problems that human vision also solves these methods can be compared with human methods, hopefully leading to insights in human visual processes.

This paper addresses a method for recovering structure (i.e., depth information) by observing the motion of objects in the image plane. The method assumes that the objects are reasonably distant from the observer, so that perspective effects are small and the objects can be treated as if they were seen in parallel projection. This process is called structure from motion, after Ullman [18,19].

Previous work in this area has been of two kinds: psychological studies investigating human ability in this process, and computational investigations studying how structure can be recovered from camera images. The psychological research is reviewed in section 4; here it will only be mentioned that human beings can recover structure purely from observing the image plane positions of points. Human beings can recover the structure of as few as two connected points moving in space.

Computational studies have investigated two kinds of objects: rigid objects with several visible points, and jointed objects with two visible points on esch rigid part. The studies of Jointed objects were stimulated by some experiments by Johansson [9-13], who showed subjects movies of people moving around dark rooms with lights on their major joints, so that there were two points on each rigid part. Subjects had no difficulty correctly interpreting the movies. These studies will be reviewed briefly here, since a more detailed discussion is available in [22].

Ullman [18,1 9] proved that it was possible to recover the structure of several points on a rigid object given several accurate two-dimensional views of the points. Roach and Aggarwal [17] obtained similar results in the case of central (perspective) projection. One problem with this method was its sensitivity to noise. Another problem was the inability to explain human ability in recovering the structure of as few as two connected points. The assumptions used by these researchers are not strong enough to allow the recovery of structure in this case.

Several attempts have been made to recover the structure of jointed objects like those studied by Johansson. Clocksin [4] and Raahid [16] used heuristic methods to recover the connectedness structure. O'Rourke and Badler [15] used a known human figure to guide recovery of three-dimensional structure. Hoffman and Flinchbaugh [7] used the assumption of motion planarity to guide recovery of both the connectedness structure and the three-dimensional structure. This planarity assumption is a special case of the assumption used here.

The present research attempts to unify the recovery of the structure of Johansson-type (jointed) figures and rigid objects. This is done by adding the fixed axis assumption: all movement consists of translations and rotations about a fixed axis. This assumption makes it possible to recover the structure of as few as two rigidly

* This research supported by the Air Force Office of Scientific Research under grant number AFOSR 77-3190.

connected points. The fixed axis assumption also gives a criterion which can be used to decide whether two points are connected, so that both the connectedness structure and the three-dimensional structure of Johansson-type figures can be recovered.

The algorithm for the recovery of the structure of rigid and jointed objects is developed in the next section. An experiment testing performance on Johansson's data is presented in section 3. Psychological implications are discussed in section 4.

2. RIGID AND JOINTED OBJECTS

An algorithm for recovering the structure of rigid and jointed objects is developed in this section. First the fixed axis assumption is justified on the basis of everyday observations of movement. Then it is proved that the fixed axis assumption makes possible the recovery of the structure of any group of rigidly connected points. The algorithm for the recovery of structure is developed at the same time. The algorithm is then extended to jointed objects. The development for jointed objects is brief because a more complete development is available in [21].

Many natural and man-made objects move in ways that tend to satisfy the fixed axis assumption. Objects that move with a fixed orientation to a plane must satisfy the assumption, because only planar rotations are possible, so that the axis must point out from the plane. The bodies of automobiles are an example of this. Objects that move through the air, such as frisbees and maple seeds, tend to satisfy the fixed axis assumption because they rotate rapidly around a fixed axis to maintain stability.

Jointed objects also often satisfy this assumption. Many animal locomotive systems use fixed axis movements extensively. For example, the walking motion of humans involves rotations about a line roughly perpendicular to the line of travel and parallel to the plane the human walks on. At the same time, the person's arms rotate about an axis parallel to this line. Many gymnastics movements consist of rotations about a fixed axis. For reasons of simplicity, much machine design is based on planar translations and rotations. For example, the wheels on a car rotate rapidly about a fixed axis.

Using the fixed axis assumption, it is shown that any rigid object satisfying it has the following property: fixing any point on the object causes the other points to trace out circles in planes normal to a fixed axis. Under parallel projection, these circles project onto ellipses in the image plane. The orientation of each circle, and hence the structure of the rigid object, can be recovered to within a reflection by finding the equations describing the ellipses.

First it is shown that fixing any point on the rigid object in space reduces the motion of any other point to a circle in a plane normal to a fixed axis. Let T be a rigid object rotating about an axis with direction $\bar{v}$, $|\bar{v}|=1$ (i.e., $\bar{v}$ is a unit vector), passing through a point O in T. Let R and P be any two feature points in T, and let $\bar{R}$, $\bar{P}$, and $\bar{O}$ be the positions of R, P, and O in some frame. The motion of P relative to R, which will be called the reference point, will be shown to be a circle in a plane normal to the axis of rotation. Let $\bar{P}-\bar{R}=a\bar{v}+\bar{\alpha}$, where $\bar{v}\cdot\bar{\alpha}=0$, and let $\bar{R}-\bar{O}=b\bar{v}+\bar{\beta}$, where $\bar{v}\cdot\bar{\beta}=0$. This breaks the position of $\bar{P}-\bar{R}$ and $\bar{R}-\bar{O}$ into two components: one along the axis of rotation and the other perpendicular to it. Similarly, let $\bar{P}-\bar{O}=c\bar{v}+\bar{\gamma}$, where $\bar{v}\cdot\bar{\gamma}=0$. Then define the function $f(\bar{x})=\bar{x}-\bar{R}$, which fixes R at the origin. The effect of the function f on the points in T is to be determined. Let T move by a translation and a rotation about $\bar{v}$. Let $\bar{P}'$, $\bar{R}'$, and $\bar{O}'$ be the new positions of R, P, and O. Since T is rotating about $\bar{v}$, it must be the case that $\bar{P}'-\bar{O}'=c\bar{v}+\bar{\delta}$ and $\bar{R}'-\bar{O}'=b\bar{v}+\bar{\varepsilon}$, for some $\bar{\delta}$ and $\bar{\varepsilon}$ such that $\bar{\delta}\cdot\bar{v}=\bar{\varepsilon}\bar{v}=0$. That is, the positions of P and R along the axis of rotation are not affected, although the positions normal to the axis of rotation may be changed. The distances from the axis of rotation are fixed, however. Then

$$\begin{aligned}\bar{P}'-\bar{R}'&=(\bar{P}'-\bar{O}')-(\bar{R}'-\bar{O}')\\&=(c-b)\bar{v}+(\bar{\delta}-\bar{\varepsilon})\end{aligned}$$

But

$$\begin{aligned}\bar{P}-\bar{R}=a\bar{v}+\bar{\alpha}&=(\bar{P}-\bar{O})-(\bar{R}-\bar{O})\\&=(c-b)\bar{v}+(\bar{\gamma}-\bar{\beta})\end{aligned}$$

Taking the dot product with $\bar{v}$,

$$\begin{aligned}(a\bar{v}+\bar{\alpha})\cdot\bar{v}&=(c-b)\bar{v}\cdot\bar{v}+(\bar{\gamma}-\bar{\beta})\cdot\bar{v}\\a&=c-b\end{aligned}$$

Hence $\bar{P}'-\bar{R}'=a\bar{v}+\bar{\omega}$, for some $\bar{\omega}$ such that $\bar{\omega}\cdot\bar{v}=0$. This shows that P lies in a plane normal to the axis of rotation, passing through the axis of rotation at $a\bar{v}$. To complete the proof, it must be shown that $|\bar{\omega}|=|\bar{\alpha}|$, since this shows that P stays a fixed distance from the axis of rotation and therefore traces out a circle around it. Now $|\bar{\omega}|=|\bar{\gamma}-\bar{\beta}|$, but $|\bar{\delta}|=|\bar{\gamma}|$, $|\bar{\varepsilon}|=|\bar{\beta}|$, and $\bar{\delta}\cdot\bar{\varepsilon}=\bar{\gamma}\cdot\bar{\beta}$ by rigidity. Hence $|\bar{\omega}|=|\bar{\delta}-\bar{\varepsilon}|=|\bar{\alpha}|$, showing that fixing R reduces the motion of P to a circle about an axis through R with direction $\bar{v}$.

Now it is shown that the image produced by fixing R then projecting T is the same as the image produced by projecting T and then subtracting the projected position of R. The projection $pr(\bar{Q})$ using parallel projection along the z axis (i.e., in direction $\bar{k}$) of a point $\bar{Q}$ is $\bar{k}\times(\bar{Q}\times\bar{k})$, so that

$$\begin{aligned}pr(f(\bar{P}))&=\bar{k}\times(f(\bar{P})\times\bar{k})\\&=\bar{k}\times[(\bar{P}-\bar{R})\times\bar{k}]\\&=pr(\bar{P})-pr(\bar{R})\end{aligned}$$

This shows that the image showing R fixed can be produced from the image which is available, which shows R moving, simply by subtracting the projected position of R.

The first step in the algorithm subtracts the observed position of any feature point, called the reference point, from the observed positions of all other feature points. Under the fixed axis assumption, this reduces the observed motion to a rotation about a fixed axis through the origin. The rest of this section assumes this process has been applied to the image. Let P be a point on the object and let v be a unit vector in the direction of the axis. The goal is to find P, the actual position of P in three-dimensional space, given pr(P), the observed position.

It is now shown that the circle traced out by the positions of any feature point relative to the reference point projects onto an ellipse. Let $\overline{P}=c\overline{v}+\overline{\omega}$, where $\overline{v}\cdot\overline{\omega}=0$ and let $|\overline{\omega}|=r$. The radius of the circle traced out by P is r, and the plane this circle lies in cuts the axis of rotation at cv. Assume $\overline{v}\neq\overline{k}$, since if $\overline{v}=\overline{k}$ the structure cannot be recovered since each image is an image plane rotation of any other image (P traces out a circle in the image plane).

Choose α, β, and γ such that

$$\overline{\omega}=\alpha\,\overline{v}x\overline{k}+\beta\,\overline{v}x(\overline{v}x\overline{k})+\gamma\,(\overline{v}x\overline{k})x[\overline{v}x(\overline{v}x\overline{k})]$$

Such α, β, and γ can always be found because the three vectors on the right-hand side form a linearly independent set, as shown in figure 3. Then

$$\overline{\omega}\cdot\overline{v}=\alpha\,\overline{v}\cdot(\overline{v}x\overline{k})+\beta\,\overline{v}\cdot[\overline{v}x(\overline{v}x\overline{k})]+\gamma\,\overline{v}\cdot\{(\overline{v}x\overline{k})x[\overline{v}x(\overline{v}x\overline{k})]\}$$
$$0=0+0+\gamma\overline{v}\cdot\{(\overline{v}x\overline{k})x[\overline{v}x(\overline{v}x\overline{k})]\}$$

But $\overline{v}\cdot\{(\overline{v}x\overline{k})x[\overline{v}x(\overline{v}x\overline{k})]\}\neq0$, since $\overline{v}\neq\overline{k}$, so that $\gamma=0$. Hence $\omega=\alpha\overline{v}x\overline{k}+\beta\overline{v}x(\overline{v}x\overline{k})$, so that

$$pr(\overline{\omega})=\alpha\overline{k}x[(\overline{v}x\overline{k})x\overline{k}]+\beta\,\overline{k}x\{[\overline{v}x(\overline{v}x\overline{k})]x\overline{k}\}$$
$$=\alpha\overline{v}x\overline{k}+\beta\,(\overline{v}\cdot\overline{k})\overline{k}x(\overline{v}x\overline{k})\qquad(1)$$

But

$$r^2=|\overline{\omega}|^2=|pr(\overline{\omega})+\overline{k}(\overline{\omega}\cdot\overline{k})|^2$$
$$=|\alpha\overline{v}x\overline{k}+\beta(\overline{v}\cdot\overline{k})\overline{k}x(\overline{v}x\overline{k})+\overline{k}[\alpha\overline{v}x\overline{k}+\beta\overline{v}x(\overline{v}x\overline{k})]\cdot\overline{k}|^2$$
$$=\alpha^2|\overline{v}x\overline{k}|^2+\beta^2(\overline{v}\cdot\overline{k})^2|\overline{k}x(\overline{v}x\overline{k})|^2+\beta^2|\overline{v}x\overline{k}|^4$$

But $|\overline{k}x(\overline{v}x\overline{k})|=|\overline{v}x\overline{k}|$ and $1=|\overline{v}|^2=(\overline{v}\cdot\overline{k})^2+|\overline{v}x\overline{k}|^2$ so

$$r^2=\alpha^2|\overline{v}x\overline{k}|+\beta^2|\overline{v}x\overline{k}|^2$$

or

$$\left[\frac{r}{|\overline{v}x\overline{k}|}\right]^2=\alpha^2+\beta^2\qquad(2)$$

Now consider a new coordinate system (p,q) defined by the orthonormal vectors $\frac{\overline{v}x\overline{k}}{|\overline{v}x\overline{k}|}$ and

kx(vxk)/ Changing the coordinate
l'vxlcl

system in this way does not affect geometric properties. Equations (1) and (2) show that in this coordinate system, the projection of w satisfies: (p,q) can be on the projection of some w only if

$$p^2+q^2(\overline{v}\cdot\overline{k})^2=r^2$$

or

$$\left[\frac{p}{r}\right]^2+\left[\frac{q}{\frac{r}{|\overline{v}\cdot\overline{k}|}}\right]^2=1$$

so that the projections of ω trace out an ellipse in the image plane. The major axis of this ellipse has length r, and the minor axis has length $\frac{r}{|\overline{v}\cdot\overline{k}|}$, so that the ratio of the lengths of the axes gives the z-coordinate of the axis of rotation (to within a sign).

Moreover, the direction of the minor axis is (0,1) in this coordinate system, which may be expressed as_kx(vxk)_ in vector notation. Since pr(p)-ckx(vxk)+kx(wxk), and since v is constant, P projects onto an ellipse with center at ckx(vxk). Thus the minor axis of the ellipse and a line from the origin to its center are collinear, and the projection of v can be recovered from either parameter, or the two can be combined to constrain the search space. Since the z-coordinate of v is known from the ratio of the lengths of the axes of the ellipse, this makes is possible to recover v. Finally, w can be found since w=0 and I W ' T .

The equations given above could be used to solve for the structure of the rigid object by fitting a non-linear equation. It is more convenient to develop the equations in a different way, so as to simplify the equation fitting process.

Let (x_{if},y_{if}) be the observed coordinates of several feature points, $1\leq i\leq k$, in several frames, $1\leq f\leq F$. If the data is rotated counter-clockwise through the angle θ by the transformation

$$\begin{aligned}u_{if}&=x_{if}\cos\theta-y_{if}\sin\theta\\v_{if}&=x_{if}\sin\theta+y_{if}\cos\theta\end{aligned}\qquad(3)$$

the ellipses will be rotated so all centers are on the u-axis. This class of ellipses is described by the equation

$$(C^2+1)u_{if}^2+v_{if}^2+D_iu_{if}=F_i\qquad(4)$$

Here C is a constant for all ellipses (C is related to the depth coordinate of the direction of the axis of rotation), and D_i and F_i may be different for each ellipse.

Note that given C and θ one can find D_i and F_i using least squares, so that equations (3) and (4) describe a non-linear system in two variables. In fitting the equations, it is more convenient for both non-linear variables to be of the same type, so that the equation fitting is actually done using C and tan θ as the non-linear variables. This method has been used to program the three-dimensional reconstruction phase of the algorithm, using Fletcher-Powell minimization as described by Acton [1].

This technique works well enough when there are only two feature points on the rigid object.

However, performance does not improve significantly when there are several points on the rigid object; the equations developed do not take advantage of useful constraints on the relative notions of the points in space. Equations (3) and (4) do not require the object to be truly rigid; each point can move independently of the others, though each must trace out a circle in a plane normal to the axis of rotation.

Now the algorithm for jointed objects will be developed. The first step in the algorithm applies the rigid part algorithm to the image, testing every point as a reference point. This divides the image into rigid parts, and determines the three-dimensional structure of each part. Next the jointed objects have to be built up out of these parts.

Joints that are seen (i.e., joints with feature points) are identifiable as joints because the feature points corresponding to them belong to more than one rigid part. Once a seen joint is found, the structure of the two rigid parts it connects must be found. The rigid part algorithm finds two reflections of each rigid part. Each reflection makes a different prediction about the position of the joint. When the two rigid parts are put together, the predicted positions must be consistent, so that each reflection of a rigid part is consistent with only one reflection of the other rigid part. This means that when the two rigid parts are put together there are only two valid reflections of the two rigid parts together. For example, in a side view of a walking human each of the four limbs consists of two rigid parts. Each rigid part has two interpretations as given by the rigid part algorithm, so that there would be 256 interpretations of the limbs of the body. However, when the predicted positions of the joints are resolved, the 256 interpretations would be reduced to 16, since each limb has two interpretations.

The jointed object can still be assembled if the Joints are not seen. Once the rigid part algorithm has analysed the structure of the rigid parts, the relative positions of the points and the direction of the axis of rotation are known. The position of the Joint (if it exists) can then be expressed relative to the reference point as a linear combination of a vector from the reference point to any other point, a vector in the direction of the axis of rotation, and the cross product of these two vectors. These positions can be resolved for any two rigid parts using least-squares minimisation of a linear form, and then the residual error can be examined to see if it is reasonable to suppose that a Joint exists between the rigid pa rts.

The algorithm for recovering structure of rigid and jointed objects begins by analysing the structure of all rigid parts in the scenes. Then the jointed objects are assembled, in two different ways depending on whether the joints are seen or unseen. The result is that the scene is analysed as a collection of rigid and jointed objects, with the structure of each being known to within a scaling factor and a number of reflections.

3. EXPERIMENTAL RESULTS

The structure from motion method developed in this paper has been implemented and run on Johansson'8 [9-13] data. These data were obtained from a commercial film made for educational purposes [8]. The data were encoded by hand using a Numonics Graphics Data Calculator. No three-dimensional information was available for this data. Two sets of data with three-dimensional information were tested in [21 J.

The data is shown in figure 1 . Six points were available: shoulder, elbow, wrist, hip, knee, and ankle. Twenty-one frames were analysed: since the film speed twenty-four frames per second the time interval was just under one second.

The connectedness structure found by thresholding the residual errors in the fits of the ellipses is shown in figure 2. The proposed structure seems perfectly reasonable, since the shoulder and hip are not a single rigid part. (including the shoulder — hip connection would have forced the shoulder — wrist and hip — ankle connections) .

The rigid part lengths proposed by the algorithm are shown in Table I. These rigid part lengths were calculated by assuming the maximal shoulder — ankle extension (as measured from the image plane view) was 4.5 feet. Most of these lengths seem reasonable; clear exceptions are the shoulder — elbow and knee — ankle lengths. The algorithm seems to be recovering part of the structure of the object, but the results are not good enough to be used for three-dimensional structure analysis.

This section has reported the results of an experiment testing the behavior of the algorithm on Johansson's data. The connectedness structure recovered is reasonable. The rigid part lengths inferred seem less so.

4. PSYCHOLOGICAL IMPLICATIONS

The previous sections have developed an algorithm that can recover structure from motion in many situations. In this section the capacity of the algorithm to explain some aspects of human motion vision is examined.

The perception of depth by humans is a complicated process Involving many independent sources of information, including stereo vision, perception of slant and tilt, and motion parallax [5]. The present research addresses the perception of structure from motion of rigid and jointed structures under conditions where parallel projection is a reasonable approximation of the camera conditions.

The perception of structure from motion under various kinds of translation and rotation has been a popular topic of psychological research in vision. Johansson [l2] clearly suggests the role projective geometry and rotation may play in this process. The first major study of this phenomenon was by Wallaoh and 0*Connall [20], who demonstrated what thay oallad tha "kinetic dapth affaot'' by showing subjects tha shadow of a rotating wire figure. Later White and Mueeer [23] demonstrated that subjects could under some conditions recover three-dimensional structure accurately.

The present algorithm will recover the correct three-dimensional structure when the objects observed do not trsnslste in depth and rotate about a fixed axis. This is consistent with human behavior as observed by several researchers and summsrised by Braunstein [3, page 107]. If there is no rotation or the axis of rotation is perpendicular to the image plane, the algorithm cannot determine structure. This is consistent with human motion vision, which sees planar motion in the second case, as observed by Borjesson and Von Hofaten [2].

Rigidity-based algorithms, like those of Ullman [18, 19] and Roach and Aggarwal [17], predict alightly different behavior: when there are only two points visible, or in somecasesa where the points observed are coplanar, the correct atructure cannot be determined. Human motion viaion does not have thia restriction, as noted in Johanna son's experiments [9-13, esp. 13] and by Borjesaon and Von Hofsten [2 J.

In some cases the algorithm predicts failure in grouping into rigid parte. For example, in the case of tumbling motion the algorithm may not succeed in finding the correct grouping, ao that objects that are actually rigid would appear to be non-rigid. Tumbling motion has been studied by Green [6], who found that rigidity is less pronounced with tumbling than with simple rotation. Evidence on thia point ia still unclear; several factors seem to influence the perception of rigidity, and tumbling motion has apparently been Investigated only by Green. Thia result might be explained by a rigidity-based algorithm if the implementation of the algorithm made fixed axia motion easier to interpret, auch as in Ullman's ([is], pp. 237-238) proposed implementation of his rigidity-based algorithm by an expert system in which each expert interpreta motion about a single axis. Further experiments would aid understanding of human motion vision under these conditione.

The experimental results reported in the last section suggest that the proposed algorithm can account for some aspects of human behavior in interpreting Johansson-type figures. It seems likely that some high-level knowledge ia used by humans in Interpreting these images. One way of using thia knowledge has been Inveatlgated by O'Rourke and Badler [15]. Another kind of higher el knowledge has been investigated by Cutting [5]. Thia research attempted to explain human ability in determining gender from impoverished Johansson-type figures (for example, subjects could determine gender with better than chance accuracy even when only the ankles were seen). It may be that the fixed axis assumption can be incorporated in a useful way with high-level type knowledge to aid interpretation of human movement. Thia synthesis could be useful not only for understanding human visual processes, but also for aiding in the interpretation of athletic movement from monocular images [14].

One kind of motion that the present algorithm cannot account for is motion Involving s large depth component. Its behavior in the presence of objects moving in depth will depend on the relative amount a of motion in depth and rotation. With a large rotation component and a fast camera, the algorithm will perform correctly. If the rotation component ia small, the algorithm will still group the points correctly, but the structure will not be recovered. The structure the algorithm assigns has the the other points swinging about the reference point. Ullsan [19] observed this interpretation in an experiment involving the motion of a cylinder in depth. If the rotation and depth motion components are similar, the algorithm will fail, which ia inconsistent with human behavior as noted by Borjesaon and Von Hofaten [2].

Thia section has summerited the psychological implications of the present approach to motion viaion. The present approach can account for several aspects of human behavior in motion viaion. In addition, the preaent theory predicts interesting things in human ability, auch as the difficulty in interpreting tumbling motion. The algorithm cannot account for the correct perception of simultaneous rotation and translation in depth.

5. SUMMARY

A method for structure from motion of rigid and jointed objecta has been preeented. The method reliea upon a motion assumption. The mathematical basis of this method has been developed. Teats done with Johansson's data have been presented. Psychological implications of this method have been discussed.

ACKNOWLEDGMENTS

The authors are indebted to aeveral persons who made contributions to thia paper; in particular we would like to acknowledge many helpful discussions with Larry Davis and Worthy Martin, both of the UT Department of Computer Sciences. The comments and proofs of Mike Brady, of the MIT AI Lab, greatly aided the mathematical analysis in thia paper. The authors also gratefully acknowledge the comments of the reviewera. Linda Melnick did the typing of early drafts.

BIBLIOGRAPHY

[1] Acton, F.S., Numerical Methoda That Work, Harper and Row, New York, 1970.

[2] Borjesaon, and C. Von Hofaten, Spatial determinants of depth perception in two-dot motion patterns, Perception and Psychophyslca 11 (1972) 263-268.

[3] Braunstein, Depth Perception Through Motion. Academic Press, New York, 1971.

[4] Clocksin, W. F., Inference of structural descriptions from visual examples of motions: preliminary results. DAI Working Paper 21, Department of Aritificial Intelligence, Univ. of Edinburgh, May, 1977.

[5] Cutting, J. E., A biomechanical invariant for gait perception, J. Exp. Psych 4, 357-372, 1978.

[6] Green, B. F., Jr., Figure coherence in the kinetic depth effect, J. Exp. Psych. 63 (1961) 272-282.

[7] Hoffman, D. D. and Flinchbaugh, B. F., The interpretation of biological motion, MIT AI Memo 608, December, 1980.

[8] Houghton-Mifflin, 2-Dimensional Motion Perception, 1971.

[9] Johansson, G., Visual perception of biological motion and a model for its analysis, Perception and Psychopysics 14 (1 973) 201 -211.

[10] Johansson, G., Spatio-temporal differentiation and integration in visual motion perception, Psych. Research 38 (1976) 379-383.

[II] Johansson, G., Visual motion perception, Scientific American (November, 1976) 76-88.

[12] Johansson, G., Visual perception of rotary motions as transformations of conic sections, Psychologla 17, 226-237, 1974.

[13] Johansson, G. and Jansson, G., Perceived rotary motion from changes in a straight line. Perception and Psychophysics 4 (1968) 1 65-1 70.

[14] Messier, S., Webb, J. A., and Aggarwal, J. K., A single camera interpretation of three-dimensional motion as an alternative to multi-camera analysis. Proc. Panamerican congress of sports medicine and exercises, May, 1981.

[15] O'Rourke, J. and Badler, N., Model-based image analysis of human motion using constraint propagation, IEEE PAMI PAMI-2 (November, 1980) 522-536.

[16] Rashid, R. F., Towards a system for the interpretation of moving light displays, IEEE PAMI PA MI-2 (November, 1980) 574-581.

[17] Roach, J. W., and Aggarwal, J. K., Determining the movement of objects from a sequence of images, IEEE PAMI PAMI-2 (November, 1980) 554-562.

[I8] Ullman, S., The interpretation of visual motion, Ph. D. thesis, Department of Electrical Engineering and Computer Science, Mass. Inst. Tech. (May, 1977;.

[19] Ullman, S., The Interpretation of Visual Motion, The MIT Press, Cambridge, Massachusetts (1979).

[20] Wallach, H. and O'Connell, D. N., The kinetic depth effect, J. Exp. Psych. 45 (1953) 205-217.

[21] Webb, J. A. and Aggarwal, J. K., Visual interpretation of the motion of objects in space, IEEE Conf. on Pattern Recognition and Image Processing, Dallas, Texas, August 1981.

[22] Webb, J. A. and Aggarwal, J. K., Visually interpreting the motion of objects in space, IEEE Computer, to appear.

[23] White, B. W., and Mueser, G. E., Accuracy in reconstructing the arrangement of elements generating kinetic depth displays, J. Exp. Psych. 60_ (July, 1960) 1-11.

From	To	Est. Length (Feet)
Shoulder	Elbow	4.2
Elbow	Wrist	0.94
Shoulder	Hip	1.5
Hip	Knee	1.7
Knee	Ankle	2.4

TABLE I. Rigid part lengths, Johansson data

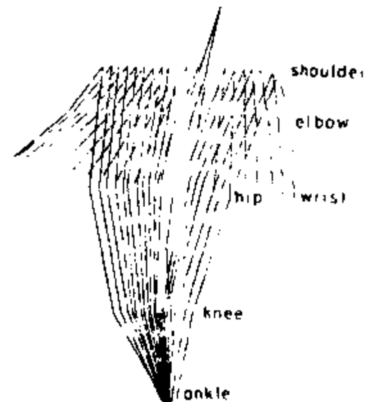

Figure 1. Johansson data. A registration occurs midway through this sequence. The present algorithm is insensitive to such errors.

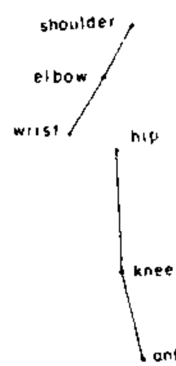

Figure 2. Connectedness structure for Johansson data.