MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL INFORMATION PROCESSING
WHITAKER COLLEGE

A.I. Memo No. 845 October, 1985
C.B.I.P. Memo No. 015

The Incremental Rigidity Scheme for Recovering Structure from Motion: Position vs. Velocity Based Formulations

Norberto M. Grzywacz and Ellen C. Hildreth

ABSTRACT: Perceptual studies suggest that the visual system uses the "rigidity" assumption to recover three-dimensional structure from motion. Ullman (1984) recently proposed a computational scheme, the *incremental rigidity scheme*, which uses the rigidity assumption to recover the structure of rigid and nonrigid objects in motion. The scheme assumes the input to be discrete positions of elements in motion, under orthographic projection. We present formulations of Ullman's method that use velocity information and perspective projection in the recovery of structure. Theoretical and computer analyses show that the velocity-based formulations provide a rough estimate of structure quickly, but are not robust over an extended time period. The stable long term recovery of structure requires disparate views of moving objects. Our analysis raises interesting questions regarding the recovery of structure from motion in the human visual system.

Acknowledgments. This report describes research done within the Artificial Intelligence Laboratory and the Center for Biological Information Processing (Whitaker College) at the Massachusetts Institute of Technology. Support for the A.I. Laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-80-C-0505. Support for this research is also provided by a grant from the Office of Naval Research, Engineering Psychology Division. NMG is supported by a Bentrell Fellowship.

Table of Contents

1. Introduction

An important source of three dimensional information is provided by the relative motions of elements in the changing two dimensional image. The human visual system is capable of recovering structure from motion, under both orthographic and perspective projection, and in the absence of all other cues to 3 D structure (see, for example, Miles, 1931; Wallach and O'Connell, 1953; Braunstein, 1976; Johansson, 1978; Ullman, 1979). In studying the computation of structure from motion, one immediately faces the problem that the recovery of structure is underconstrained; there are infinitely many 3 D structures consistent with a given pattern of motion in the changing 2 D image. Additional constraint is required to establish a unique interpretation.

Early perceptual studies suggested that the rigidity of objects may play a key role in the recovery of structure from motion (Wallach and O'Connell, 1953; Gibson and Gibson, 1957; Green, 1961; Johansson, 1975, 1977). Computational studies later established that rigidity is a sufficiently powerful constraint to derive a unique interpretation of structure, under a variety of viewing conditions. For example, Ullman and Fremlin (Ullman, 1979) showed that under orthographic projection, three views of four non coplanar points are sufficient to guarantee a unique 3 D interpretation (up to an unavoidable reflection about the image plane). In the case of perspective projection, Longuet Higgins and Prazdny (1981) proved that the instantaneous velocity field and its first and second spatial derivatives at a point admit at most three different 3 D interpretations. Tsai and Huang (1981) showed that, with the exception of a few special configurations, two perspective views of seven points in motion are sufficient to guarantee a unique 3 D interpretation. Waxman and Ullman (1984) also addressed the uniqueness of the recovery of structure under perspective projection, basing their results on a kinematic analysis of continuous image flows. Additional theoretical results have been obtained for various classes of restricted motion, such as planar surfaces in motion (Hay, 1966; Longuet Higgins, 1984; Waxman and Ullman, 1985; Ullman, 1985; Negahdaripour and Horn, 1985), pure translatory motion (Clocksin, 1980), planar or fixed axis rotation (Hoffman and Flinchbaugh, 1982; Webb and Aggarwal, 1981; Bobick, 1983; Bennett and Hoffman, 1984a,b; Sugie and Inagaki, 1984), and translation perpendicular to the rotation axis (Longuet Higgins, 1983). A review of the theoretical results regarding the recovery of structure from motion can be found in Ullman (1983).

From theoretical studies of the structure from motion problem, it can be concluded that by exploiting a rigidity constraint, 3 D structure can be recovered from motion alone, using image information that is integrated over a small extent in space and in time. These theoretical studies have also given rise to algorithms for deriving the rigid 3 D structure of moving objects (for example, Ullman, 1979; Longuet Higgins, 1981; Tsai and Huang, 1981). Experimentation with these algorithms has revealed two important limitations. First, although it is possible in theory to recover structure from motion information that is integrated over a small extent in space and time, such a strategy may not be robust in practice. A small amount of error in the image measurements can lead to very different solutions (Ullman, 1983). Second, most previous algorithms derive a three dimensional structure only when a rigid interpretation is possible, and

otherwise do not yield any interpretation of structure or yield a solution that is incorrect or unstable.

The first observation above suggests that a robust algorithm for recovering structure should use motion information that is more extended in space or time. This conclusion is supported in recent computational studies by Nagahdaripour and Horn (1985) and Ullman (1984). Negahdaripour and Horn addressed the recovery of the motion of an observer relative to a stationary planar surface. It was shown that a robust recovery of both the observer motion and the orientation of the plane is possible when dense measurements of the spatial and temporal derivatives of image brightness are integrated over a large region of the changing image. Thus, consideration of motion information that is more extended in space can lead to a stable recovery of structure. Bruss and Horn (1983) also proposed an algorithm for recovering the motion of an observer relative to a stationary scene, which integrates motion information over an extended region of the image. The study by Ullman (1984), which will be developed further in this paper, demonstrated that a robust recovery of structure is also possible when motion information is integrated over an extended period of time. The extension in time can be achieved, for example, by considering a large number of discrete frames or by observing continuous motion over a significant temporal extent.

With regard to the human visual system, the dependence of perceived structure on the spatial and temporal extent of the viewed motion has not yet been studied systematically, but the following informal observations have been made. Regarding spatial extent, two or three points undergoing relative motion are sufficient to elicit a perception of 3–D structure (Borjesson and von Hofsten, 1973; Johansson, 1975), although theoretically the recovery of structure is less constrained for two points in motion, and perceptually the sensation of structure is weaker. An increase in the number of moving elements in view appears to have little affect on the quality of perceived structure (for example, Petersik, 1980). Regarding the temporal extent of viewed motion, Johansson (1975) showed that a brief observation of patterns of moving lights generated by human figures moving in the dark (commonly referred to as biological motion displays) can lead to a perception of the 3–D motion and structure of the figures. Other perceptual studies indicate that the human visual system requires an extended time period to reach an accurate perception of 3–D structure (Wallach and O'Connell, 1953; White and Mueser, 1960; Green, 1961). A brief observation of a moving pattern sometimes yields an impression of structure that is "flatter" than the true structure of the moving object (Ullman, 1984). Thus, the human visual system is capable of deriving some sense of structure from motion information that is integrated over a small extent in space and time. An accurate perception of structure may, however, require a more extended viewing period.

It was noted earlier that most algorithms for recovering structure from motion are unable to interpret nonrigid motions. There are, however, some exceptions to this. Bennett and Hoffman (1984b) studied the minimum amount of motion information required to derive a unique interpretation of the structure of a set of discrete elements undergoing nonrigid motion, when it is assumed that the elements are rotating about a fixed axis in space. Hoffman and Flinchbaugh (1982) proposed an algorithm for interpreting the 3–D motion and structure in biological motion displays. This algorithm decomposes the

overall nonrigid motion into pairs of points that are rigidly linked and rotating in a plane, and triplets of points forming two hinged links that rotate in the same plane. Koenderink and Van Doorn (1984; Koenderink, 1984) examined the class of bending deformations, which satisfy the physical constraint that distances along the surface of the object are preserved by the transformation. This class of deformations excludes any stretching or compressing of the object surface. In its current formulation, the method proposed by Koenderink and Van Doorn for recovering the structure of bending surfaces requires that the surfaces be complete, in contrast with other algorithms that are able to interpret the structure of isolated points in motion. To conclude, the algorithms discussed above for recovering the structure of nonrigid objects in motion all address restricted classes of these motions, such as fixed axis motion, planar motion, and bending deformations.

The mechanism for recovering structure from motion in the human visual system appears not to be based strictly on the rigidity assumption. It is an everyday experience to perceive the structure and motion of deforming objects such as a flowing river, an expanding balloon, or a dancing ballerina. Perceptual studies reveal that the human visual system can derive some sense of structure for a broad range of nonrigid motions, including stretching, bending and even more complex types of deformations (Johansson, 1964, 1978; Jansson and Johansson, 1973; Todd, 1982, 1984). It is also the case that displays of rigid objects in motion sometimes give rise to the perception of somewhat distorting objects (Wallach, Weisz and Adams, 1956; White and Mueser, 1960; Green, 1961; Braunstein, 1962; Sperling *et al.*, 1983; Hildreth, 1984a; Adelson, 1985).

In this paper, we focus on the recent work of Ullman (1984), which provides a more flexible method for deriving the structure of rigid and nonrigid objects in motion, and provides a natural means for integrating motion information over an extended time period. This method makes use of the rigidity assumption, but in a different way from previous studies. The algorithm, called the *incremental rigidity scheme*, maintains an internal model of the structure of a moving object, which is continually updated as new positions of image elements are considered. The initial model may be flat, if no other cues to 3-D structure are present, or it may be determined by other cues available, for example, from binocular stereo, shading, texture, and perspective (Marr, 1982; Ballard and Brown, 1982; Horn, 1985). As each new view of the moving object appears, the algorithm computes a new set of 3-D coordinates for points on the object, which maximizes the rigidity in the tranformation from the current model to the new positions. In particular, the algorithm minimizes the change in the 3-D distances between points in the model. The formulation presented by Ullman assumes the input to the recovery process to consist of a sequence of discrete frames, each containing a set of discrete feature points. Through the process of repeatedly considering a new frame in the sequence and updating the current model of the structure of the moving features, the incremental rigidity scheme builds up and maintains a 3-D model, and can be applied to both rigid and nonrigid objects in motion. Further details of the incremental rigidity scheme are presented in section 2 and in Appendix A.

The incremental rigidity scheme has a number of advantages, from a computational perspective (Ullman, 1984): (1) because it integrates information over an extended time period, it provides a stable recovery of structure, particularly in the presence of error in

the image measurements, (2) it allows deviations from rigidity, while always maintaining some model of the 3 D structure of the object, (3) it provides a natural means for interactions with other sources of 3 D information, and (4) empirical studies suggest that the algorithm is able to recover the correct 3 D structure. The behavior of this algorithm is also consistent in several ways with human perceptual behavior (Ullman, 1984).

In this paper, we develop Ullman's work further, in several directions. First, in section 2, we present a continuous formulation of the incremental rigidity scheme, which uses velocity information at discrete points as input to the recovery process. In section 3, we then examine in more detail, the behavior of the incremental rigidity scheme when presented with rigid objects undergoing rotation about a single axis in space, under orthographic projection. In particular, through computer simulations and a theoretical analysis, we examine the behavior of the discrete formulation as a function of the angular displacement between frames, and compare this behavior with that of the continuous formulation. Finally, in section 4, we present both discrete and continuous formulations of the incremental rigidity scheme that use perspective projection. Through computer simulations, we begin to examine the behavior of the perspective formulations, when presented with rigid objects undergoing both pure rotation about a single axis, and pure translation through space.

The main conclusions of the paper are the following. The direct use of velocity information as input to the incremental rigidity scheme can provide a rough estimate of the structure of a moving object over a short viewing period, but is not sufficiently powerful to allow a detailed and robust recovery of structure over an extended time period. The computation of a stable long term solution appears to require the use of views of a moving object that differ significantly. This implies the need for a recovery process with memory of the past views, but this memory need not be extended indefinitely and continuously into the past. A small number of discrete views of a moving object are sufficient for recovering 3 D structure. In the case of the incremental rigidity scheme, the use at every instant of a current model of the 3 D structure of the object, and a present view that is sufficiently different from preceding views, can provide a robust recovery of structure. In the case of rotation of a rigid object about a single axis in space, both the rate of convergence of the algorithm to the final solution and the quality of the solution decrease as smaller angular displacements between viewed frames are considered. In the limit of the continuous formulation, the solution is no longer stable. The behavior of the perspective formulation of the incremental rigidity scheme is more complex than that of the orthographic formulation. We found that if the absolute position of an object in space is known throughout the motion of the object, then the perspective formulation performs well, similar to the orthographic formulation. The results of computer simulations revealed a degradation in performance with smaller angular and spatial displacements between frames, but this degradation was somewhat more severe than in the orthographic formulation. Again, in the limit of the continuous formulation, the solution is no longer stable. Our analysis raises important questions regarding the quantitative ability with which the human visual system can recover structure from motion; these questions are discussed in section 5.

2. Discrete and Continuous Formulations of the Incremental Rigidity Scheme

In this section, we first describe Ullman's (1984) original formulation of the incremental rigidity scheme, which assumes the visual input to consist of a sequence of frames, each containing a number of discrete points that may correspond to identifiable features in the changing image. We then present a formulation that uses velocity information at discrete points in a continuously changing image as input to the recovery process. The analysis in this section assumes orthographic projection of the scene onto the image plane.

The motivations for considering a continuous formulation are threefold. First, on the basis of the results of computer simulations, Ullman (1984) noted that when analyzing objects undergoing rigid rotation, the convergence of the incremental rigidity scheme to the correct solution was slower when smaller angular separations between frames were used. This suggests that the scheme may perform better when successive views of the object differ significantly. We considered the limit of arbitrarily close frames, both as a means of studying this phenomenon, and to determine whether a robust recovery of structure is still possible under these conditions. A second motivation is that recent work on the computation of an instantaneous 2 D velocity field from the changing image suggests that a unique velocity field can be obtained for general classes of motion, exploiting a constraint on the smoothness of the velocity field (Horn and Schunck, 1981; Hildreth, 1984a,b; Nagel, 1984). Ultimately, it may be useful to integrate the results of such velocity field computations with the recovery of structure from motion. A third motivation is that Ullman's formulation of the incremental rigidity scheme leads to the solution of a set of nonlinear equations. It is shown in Appendix A that the continuous formulation presented here leads to the solution of a set of linear equations. This makes a theoretical analysis of the solution more accessible, and could in principle result in a more efficient computer implementation.

2.1 Ullman's Discrete Formulation

The incremental rigidity scheme maintains and updates an internal model $M(t)$ of the viewed object, which consists of a set of 3-D coordinates: $M(t) = (x_i(t), y_i(t), z_i(t))$. In this section, we assume orthographic projection (the case of perspective projection is addressed in section 4) onto the $X - Y$ image plane, so that $(x_i(t), y_i(t))$ are the image coordinates of the i th point, and $z_i(t)$ is the current estimate of the depth at the i th point. (We assume a left handed coordinate system, with the positive z axis pointing away from the observer.) In Ullman's formulation, when no other 3 D cues are present, the initial model $M(t)$ at $t = 0$ is taken to be flat; that is, $z_i(0) = 0$ (or some other constant value) for $i = 1, \ldots, n$, where n is the number of points in motion. In principle, other initial configurations could also be considered. The theoretical analysis of section 3 examines the long term stability of the incremental rigidity scheme, independent of the initial model of the structure of the moving points.

Given a current model $M(t)$ at time t, and the image of the moving points in a new frame at a later time t', the problem is to compute a new model $M(t')$ such that

the transformation from $M(t)$ to $M(t')$ is as rigid as possible. Since $x_i(t')$ and $y_i(t')$ are known, this requires the computation of the unknown depth values $z_i(t')$. (It is assumed that the correspondence between points in the two successive frames is known.) The new depth values are computed as follows. Let $l_{ij}(t)$ denote the distance between points i and j at time t. To make the transformation as rigid as possible, the values $z_i(t')$ for the new model are chosen so as to make $l_{ij}(t)$ and $l_{ij}(t')$ as similar as possible. For this purpose, Ullman defined a measure of the difference between $l_{ij}(t)$ and $l_{ij}(t')$ as:

$$d(l_{ij}(t), l_{ij}(t')) = \frac{(l_{ij}(t) - l_{ij}(t'))^2}{l_{ij}^3(t)}, \tag{2.1}$$

and formulated the recovery of structure as the computation of $z_i(t')$ that minimize the following overall deviation from rigidity:

$$D(t, t') = \sum_{i,j} d(l_{ij}(t), l_{ij}(t')). \tag{2.2}$$

After the values $z_i(t')$ have been determined using this minimization process, the new model $M(t') = (x_i(t'), y_i(t'), z_i(t'))$ becomes the current model. A new frame is then registered and the process repeats itself. In this way, the scheme maintains rigidity by keeping the total distances between points in the model as constant as possible. The motivation for the cubic factor in the denominator of Eq. (2.1) is that the nearest neighbors to a given point are more likely to belong to the same object than distant neighbors, so that a point is more likely to move rigidly with its nearest neighbors. The $l_{ij}^3(t)$ factor diminishes the influence of distant points in the recovery process.

It should be noted that in the case of orthographic projection, only relative depth values, $z_i(t) - z_j(t)$, can be recovered, rather than absolute depth values, because under this form of projection, the image of a given object does not change with its absolute depth. In addition, 3-D structure is determined only up to a reflection about the image plane, since the orthographic projection of a rotating object, and its mirror image rotating in the opposite direction, coincide.

2.2 The Continuous Formulation

A continuous formulation, which uses velocity information at discrete feature points, can be developed as follows. Assume again that there always exists an internal model $M(t) = (x_i(t), y_i(t), z_i(t))$. Assume also that the image velocities $\dot{x}_i(t)$ and $\dot{y}_i(t)$ are known. The problem is then formulated as the computation of the z components of velocity, $\dot{z}_i(t)$, that minimize the total continuous change in the distances between the points. The general form of the measure of overall deviation from rigidity is given by:

$$D_c(t) = \sum_{i,j} d_c(l_{ij}(t)). \tag{2.3}$$

In our analysis, we consider different possibilities for the measure, $d_c(l_{ij}(t))$. The theoretical development of section 3.2, for example, considers the behavior of the incremental

rigidity scheme, as the frames become infinitesimally close, using the following discrete measure of the change in the distance between pairs of points:

$$d(l_{ij}(t), l_{ij}(t')) = (l_{ij}^2(t) - l_{ij}^2(t'))^2. \tag{2.4}$$

The derivation of the continuous measure of rigidity is outlined in section 3.2.1. It results in the following expression for $D_c(t)$, as a function of the coordinates and velocities of the points (the arguments, t, have been omitted for simplicity):

$$D_c(t) = \sum_{i,j} ((x_i - x_j)(\dot{x}_i - \dot{x}_j) + (y_i - y_j)(\dot{y}_i - \dot{y}_j) + (z_i - z_j)(\dot{z}_i - \dot{z}_j))^2. \tag{2.5}$$

This particular measure of rigidity was used in the theoretical study for analytic simplicity. The computer simulations of the continuous formulation use the measure of change in the distance between points given again by the limit of $(l_{ij}(t) - l_{ij}(t'))^2$, for infinitesimally close frames, which is:

$$d_c(l_{ij}(t)) = (\dot{l}_{ij}(t))^2. \tag{2.6}$$

In terms of the coordinates and velocities of the points, this yields the following overall measure of deviation from rigidity:

$$D_c(t) = \sum_{i,j} \frac{((x_i - x_j)(\dot{x}_i - \dot{x}_j) + (y_i - y_j)(\dot{y}_i - \dot{y}_j) + (z_i - z_j)(\dot{z}_i - \dot{z}_j))^2}{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}. \tag{2.7}$$

(Eqs. (2.5) and (2.7) use slightly different measures of rigidity, but serve the same role as measures of overall changes of rigidity for the continuous formulation. In the present paper we do not use different notations for these measures, as it will be clear from the context which measure is used. More specifically, Eq. (2.5) is used in the theoretical analysis of section 3.2, and Eq. (2.7) is used in the computer simulations of section 3.1.) In other respects, the continuous formulation is similar to the discrete formulation. A model of the structure of the moving points is built up by continually taking into account new velocity information over an extended time period. Again, because orthographic projection is used, only relative velocities, $\dot{z}_i(t) - \dot{z}_j(t)$, can be recovered. This can clearly be seen in Eqs. (2.5) and (2.7), in which the coordinates of the points and their time derivatives all appear in differences between pairs. Further details of the continuous formulation are presented in section 3 and Appendix A.

The analyses presented in this paper mainly consider single rigid objects in motion, which are compact in the sense that the internal distances between pairs of points do not differ much from one another. In this case, the additional l_{ij}^3 factor of Eq. (2.1) has little influence on the behavior of the algorithm, so we omitted it in our theoretical analysis and computer simulations for the sake of simplicity. In general, however, a proper weighting (and not necessarily a cubic factor) of the influence of different distances among points is necessary for a better performance of the algorithm.

3. Positions vs. Velocities as Input to the Recovery of Structure

We stated earlier that on the basis of computer simulations, Ullman (1984) observed that when analyzing objects undergoing rigid rotation, the convergence of the incremental rigidity scheme to the correct solution was slower when smaller angular displacements between frames were used. In this section, we analyze this phenomenon from a theoretical perspective, focusing on the long term stability of the computed 3 D model. We first examine the behavior of the continuous formulation, which uses a current model and measured image velocities at discrete points as input to the recovery process. We then turn to the discrete formulation, which uses the discrete positions of discrete points as input, and examine the long term stability of its solution as a function of the angular displacement between frames. Our main conclusions are the following. First, for the particular class of motions considered, the discrete formulation always yields a 3 D solution that converges asymptotically to the correct solution, but the rate of convergence varies with the angular displacement. The rate of convergence increases with increasing angular displacement up to a maximum, and then decreases with further increases in this displacement. The position of this maximum depends on such factors as the type of motion and geometric structure of the points.

Although the orthographic projection is in general not physically valid, it is used here because it allows a simpler formulation of the problem, and is therefore better suited to theoretical analysis and computer implementation. It allowed us to gain a deeper insight into the nature of the phenomena studied. We nevertheless implemented the equivalent perspective formulation and confirmed that the basic results remain valid under this formulation. The use of perspective projection enables the recovery of the structure of objects undergoing pure translation, which was not possible under the orthographic projection. In the case of translation, the rate of convergence of the computed 3 D model to the true structure also increases with increasing spatial displacements between frames.

Before presenting the results of the theoretical analysis, we illustrate the behavior of the discrete and continuous formulations through the results of computer simulations. We show that the continuous formulation yields an initial fast convergence to a close approximation of the true structure of the moving points, but then oscillates over a large range. It consequently does not yield a stable long-term recovery of structure.

3.1 Observations from Computer Simulations

In this section, we briefly illustrate the behavior of the discrete and continuous formulations of the incremental rigidity scheme, for the special case of rigid rotation of a small set of discrete points about the vertical axis. (For details of the computer implementation, see Appendix A.) In the case of the discrete formulation, we examine the rate of convergence of the algorithm and the quality of the solution that it yields, as a function of the angular displacement between frames. We then compare its performance with that of the continuous formulation. In all of the examples presented here, the input consisted of a set of five points in space. The first point is assumed to lie at the origin of a coordinate system that is displaced from the viewer along the line of sight. The position

of this first point is constant throughout the motion of the points. The coordinates of the remaining four points were chosen randomly. Fig. 1 illustrates a typical set of five points, showing their projections onto the $X - Y$ plane (Fig. 1a) and the $X - Z$ plane (Fig. 1b). In the simulations, the projected positions and velocities of the points were computed analytically, rather than measured from real image sequences.

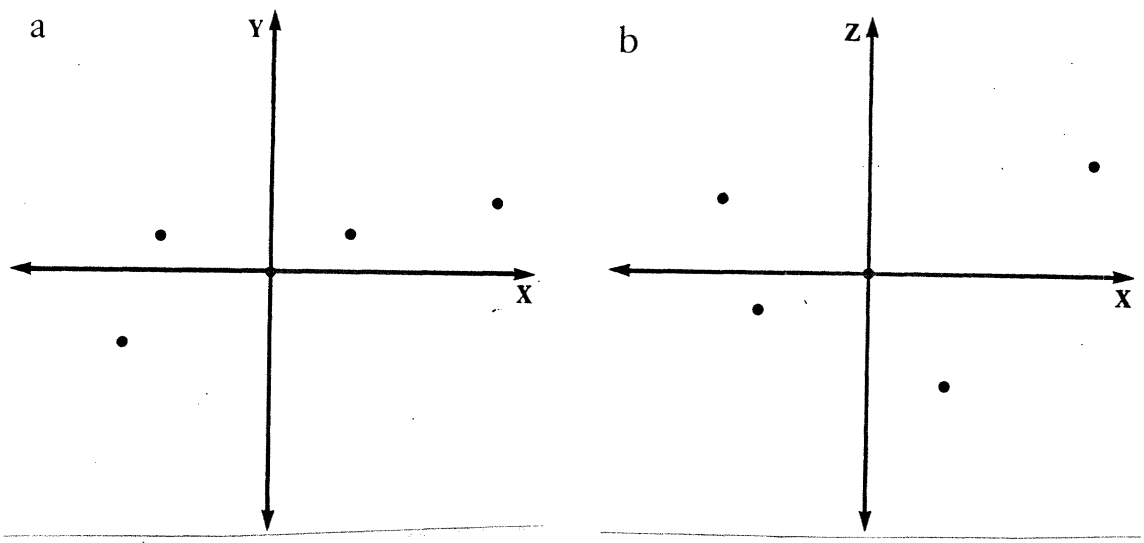

Figure 1. A set of five points with random coordinates is projected onto (a) the $X - Y$ plane, and (b) the $X - Z$ plane.

Fig. 2 illustrates the behavior of the discrete formulation of the incremental rigidity scheme, as a function of the angular displacement between frames. Each figure shows a birds' eye view of the set of rotating points (that is, their projection onto the $X - Z$ plane), with filled circles representing the true positions of the points and open circles representing the structure computed by the algorithm. Fig. 2a shows the true positions of the points and the initial model at time $t = 0$. The initial model is assumed to be flat. Figs. 2b and 2c show the true and computed configurations of points after 120° of rotation. This final position was reached by taking three steps of 40° (Fig. 2b), and 12 steps of 10° (Fig. 2c). It can be seen that the use of a smaller number of more disparate views yields a 3-D model that is closer to the true solution. As noted in Appendix A, a steepest descent minimization algorithm was used for most of our computer simulations. We also analyzed a small number of examples using an exhaustive search algorithm to find the minimum solution, and again found the accuracy of the solution to vary with the angular displacement between frames. We therefore believe this to be a fundamental behavior of the incremental rigidity scheme that is not simply a consequence of the particular algorithm used to implement the scheme. This observation is supported further by the theoretical analysis of the next section.

In Fig. 3, we show a series of graphs that illustrate in a different way, the behavior of the discrete formulation of the incremental rigidity scheme as a function of angular displacement. In this case, the set of points shown in Figs. 1 and 2 were rotated by

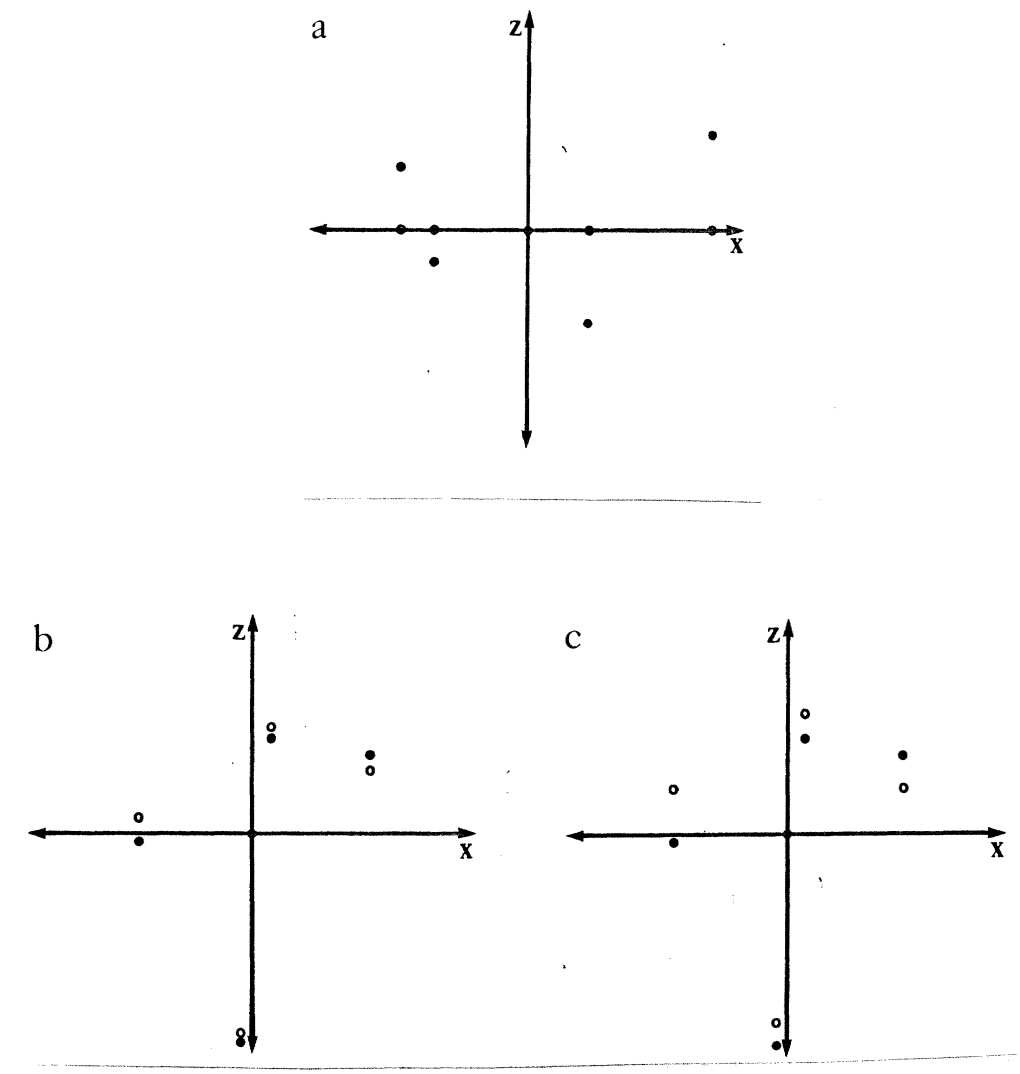

Figure 2. (a) The true configuration of five points (filled circles) is compared with the initial configuration of the points in the model (open circles) at time $t = 0$. The projection is onto the $X - Z$ plane. (b) The comparison between the true and computed positions of the points after three steps of 40°. (c) The comparison between the true and computed positions of the points after 12 steps of 10°.

four full revolutions. Each of the graphs show the error between the true and computed structures, as a function of time. In particular, the following quantity is plotted:

$$\left[\sum_{i,j}(d_{ij} - l_{ij})^2\right]^{\frac{1}{2}} \tag{3.1}$$

where d_{ij} is the correct 3 D distance between points i and j in the object, and l_{ij} is

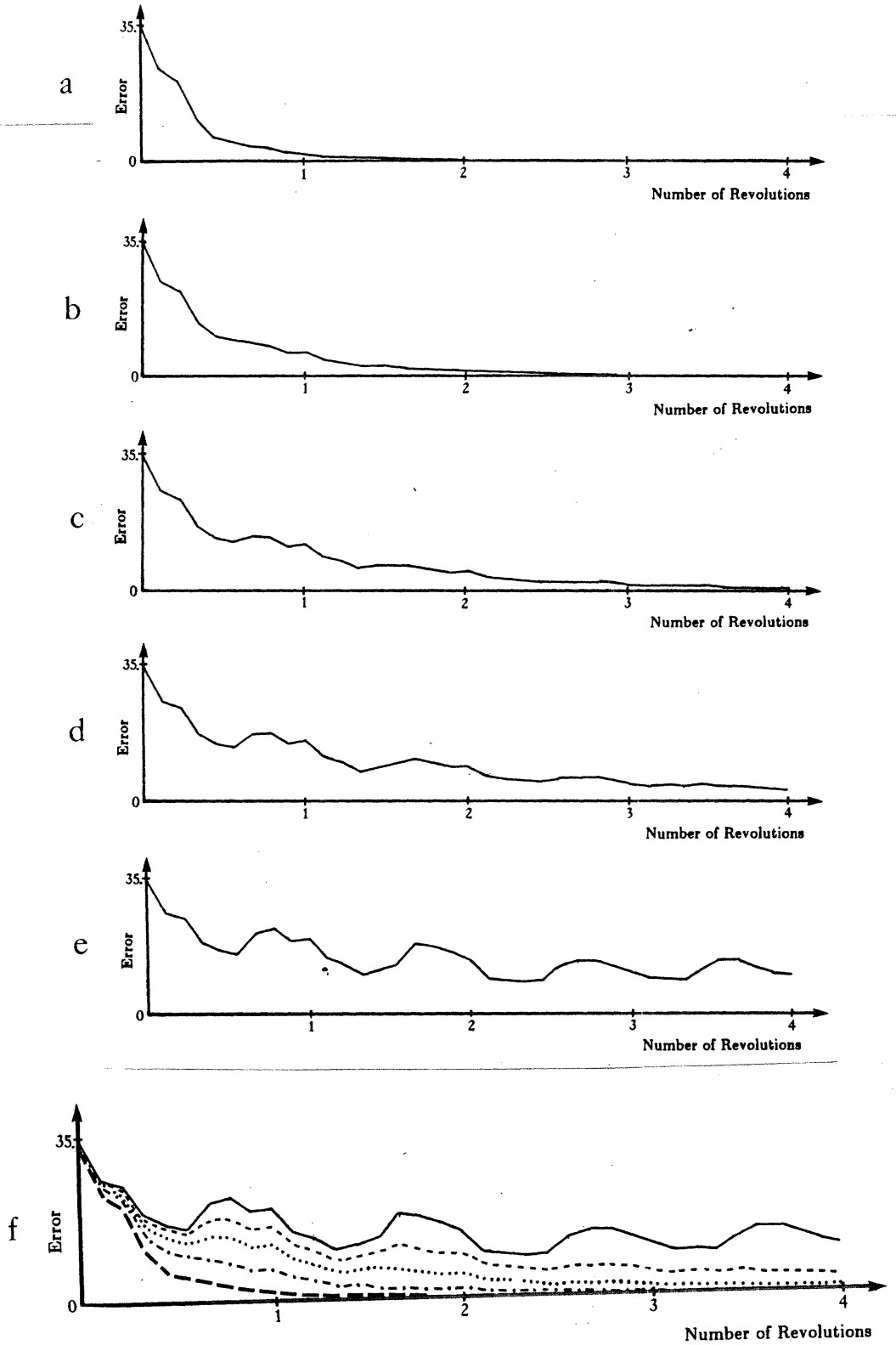

Figure 3. Graphs of the error in the internal distances between points in the computed 3-D model, as a function of time. The points are rotated 4 full revolutions, in steps of (a) 40°, (b) 20°, (c) 10°, (d) 5° and (e) 1°. (f) The graphs in 3a–3e are shown superimposed.

the corresponding distance between points i and j in the computed model. The graphs shown in Figs. 3a through 3e correspond to rotations with angular displacements of 40°, 20°, 10°, 5° and 1°, respectively. In Fig. 3f, the five graphs are shown superimposed. Again, it can be seen that the rate of convergence and quality of the solution improves with larger angular displacements. For a particular total angular extent, the error in the computed model decreases with increasing angular displacement between frames. In the case of 40° displacements, the algorithm converges asymptotically and monotonically to the final solution. For smaller displacements, the convergence is no longer strictly monotonic, but is still essentially asymptotic toward the final solution. Fig. 4 shows the same set of graphs superimposed, but with the error plotted on a log scale. The convergence of the solution is now essentially linear, with varying slopes, suggesting that the actual convergence is exponential.

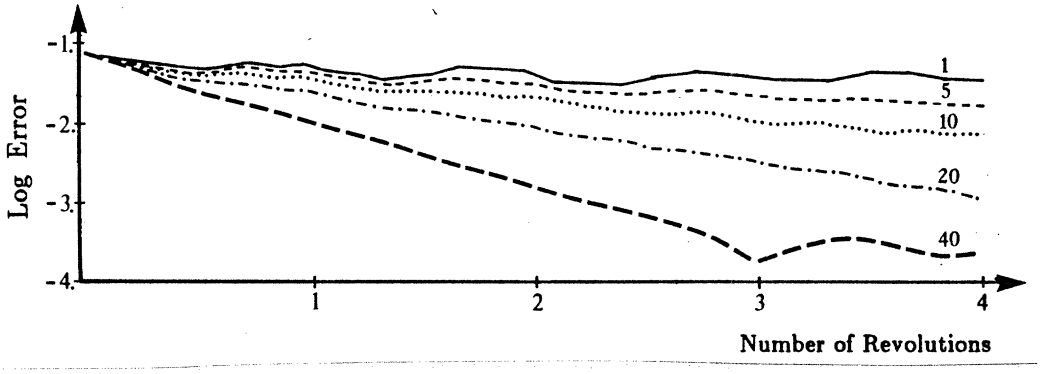

Figure 4. Graphs of the error in the internal distances between points in the computed 3-D model, as a function of time, plotted on a log scale. The points are rotated 4 full revolutions, in steps of 40°, 20°, 10°, 5° and 1°. The graphs are shown superimposed, with the angular displacements indicated above each graph.

Fig. 5 illustrates the behavior of the continuous formulation of the incremental rigidity scheme. The same set of points used previously was again rotated about the vertical axis and the 3-D model was computed at infinitesimally closely spaced times, using the instantaneous velocities projected onto the image (see Appendix A for details). In Fig. 5a, we compare the true positions of the points (filled circles) with the best solution (open circles) obtained over 10 full revolutions of the points. Although the model is quite close to the true structure at this position of the points, the solution oscillates significantly over an extended time period. A graph of the error in the computed model over the 10 revolutions is shown in Fig. 5b. There is an initial fast convergence toward the true structure of the points, but the algorithm then oscillates with high amplitude away from the true solution. When the error in the model is high, depth reversals often occur. Fig. 5c shows an example of the true and computed structures at a time of complete depth reversal. Such reversals were also observed with the discrete formulation of the scheme, although rarely.

Fig. 6 illustrates the typical behavior of the discrete formulation of the incremental rigidity scheme, averaged over 10 configurations of five points. For each of the configura-

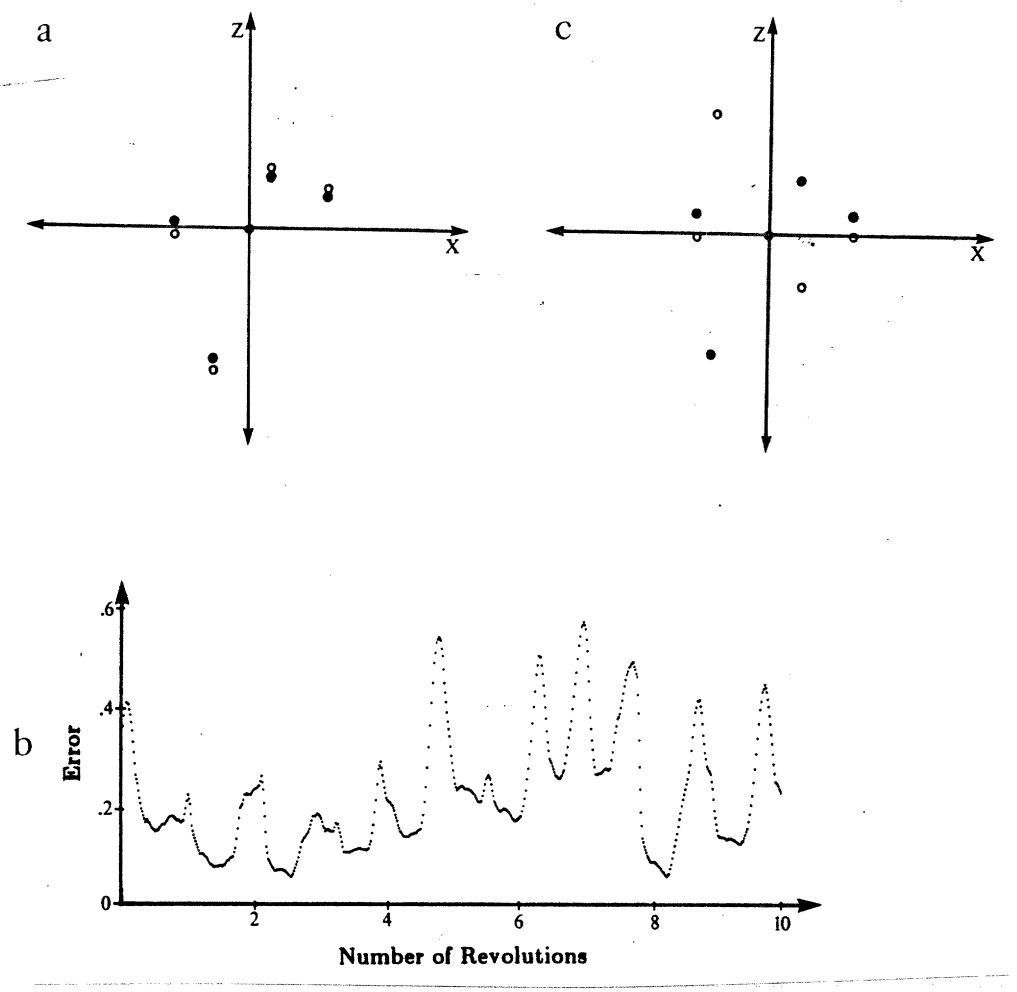

Figure 5. (a) The best solution (open circles) obtained over 10 revolutions of the points is compared with the true positions (filled circles), for the case of the continuous formulation. (b) The error in the internal 3-D distances between points in the model, as a function of time. (c) A complete depth reversal between the true and computed structures.

tions, the first point was placed at the origin of the coordinate system and the coordinates of the remaining four points were chosen randomly. The set of points was then rotated by three discrete angular steps, with the size of the angular displacements ranging from 1° to 90°. The initial 3 D model for the points was flat, and a new model was computed for each of the three discrete positions of the points. After the third step, we computed the following measure of the absolute error in the internal distances between points:

$$\frac{1}{n}\sum_{i,j}\frac{|d_{ij} - l_{ij}|}{d_{ij}} \tag{3.2}$$

where d_{ij} is the true 3 D distance between points i and j, l_{ij} is the distance between points i and j in the computed model and n is the total number of pairs of points. This measure was chosen because it expresses an average of the error in the model relative to the true structure. (Note that a lower value for this measure corresponds to less error in the computed model.) We also considered other measures of error in the distances between points and in the actual depth values, and found the general behavior of the algorithm to be the same under different measures. The above error measure was averaged over the 10 configurations of points, and is plotted in Fig. 6, as a function of angular displacement. It can first be seen that in general, the error after three discrete steps of the algorithm varies with the size of the angular displacement between frames. There is a steady improvement in performance as the displacement increases, to about 50°, followed by a degredation for increments of 60°, and a steady decrease in performance from 70° to 90°. The degradation in performance for an angular displacement of 60° was common; 8 of the 10 configurations of points exhibited this behavior. This degradation may be a consequence of the symmetry between the initial and final views, which are rotated 180° from one another. In general, the convergence of the algorithm for three discrete steps degrades significantly for smaller angular displacements. This result is not surprizing, in that there is very little change in the discrete views for such small angles and a reduced total angular extent. The deterioration for large angles probably occurs because at 90°, the number of views available to the structure from motion process is reduced to two.

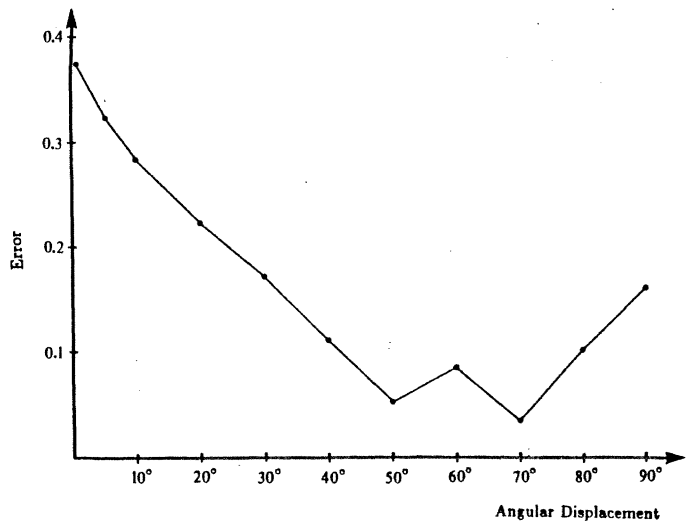

Figure 6. Absolute error in the internal distances between points in the 3 D model computed by the incremental rigidity scheme. The graph shows the error after three discrete views of the rotating points are considered, averaged over 10 random configurations of five points.

Fig. 7 illustrates the error in the computed model, as a function of angular displacement, for the case in which the overall rotation of the points was kept essentially constant. The same set of 10 random configurations of points was rotated by discrete angular steps, with the size of the steps varying from 10° to 90°. In this case, each configuration was rotated by a total of (approximately) 180° and 360°, and the same measure shown in Eq. (3.2) was then computed. This measure was again averaged over the 10 configurations of points, and is plotted as a function of angular displacement in Fig. 7. Fig. 7a shows the data for the case of 180° of rotation. Note that for the angles 40°, 50°, 70° and 80°, there are two points plotted, corresponding to multiples of the angular displacement that are just less than and greater than 180°. The graph that is superimposed on the points passes between the pairs of points for these angles. Fig. 7b shows the same data for the case of 360° of rotation (for the case of angular displacements of 50° and 70°, the points were rotated by a total of 350°). Also shown in Figs. 7a and 7b are the average errors in the solution that were derived by the continuous formulation of the algorithm after 180° and 360° of rotation of the 10 configurations of points. These two data points are indicated by the stars along the ordinate of the graph. The main observation to be made is that when the points are rotated by a constant total amount, there is still a strong dependence of the rate of convergence of the solution on the size of the angular displacement between frames. There is again an improvement in convergence rate as this angle increases, up to about 50°, followed by a slight worsening for an angle of 60°, then improvement again for 70°, followed by a steady worsening to 90°. The degradation in performance for an angular displacement of 60° was again common, occuring for 7 of the 10 configurations of points. The deterioration of the convergence rate with decreasing angular displacements is not obvious, because in spite of the fact that the changes between consecutive frames are smaller, there are many more frames altogether. In addition, while the continuous formulation provides a good estimate of the true solution after 180° (see Fig. 7a), the solution then degrades significantly, providing a relatively poor solution after 360° of rotation (see Fig. 7b).

To conclude, the results of computer experiments with the discrete formulation of the incremental rigidity scheme show a clear dependence of the behavior of the computed solution on the size of the angular displacements between frames. In the limiting case of the continuous formulation, there is an initial fast convergence of the solution toward the true solution, followed by a substantial oscillation of the solution. The next section presents a theoretical analysis that attempts to explain this phenomenon.

3.2 Analytic Study of Convergence Properties

In this section we outline the main conclusions of our theoretical analysis. The purpose of this analysis is not a general study of the behavior of the incremental rigidity scheme, as such an analysis would be too cumbersome. Rather it concentrates on a formal analysis of the convergence properties of the algorithm, for a family of examples that we believe are relevant to the general recovery of structure from motion. We emphasize the concepts raised by this analysis, and comment on their generality. This section considers a subset of those examples analyzed in the computer simulations of section 3.1.

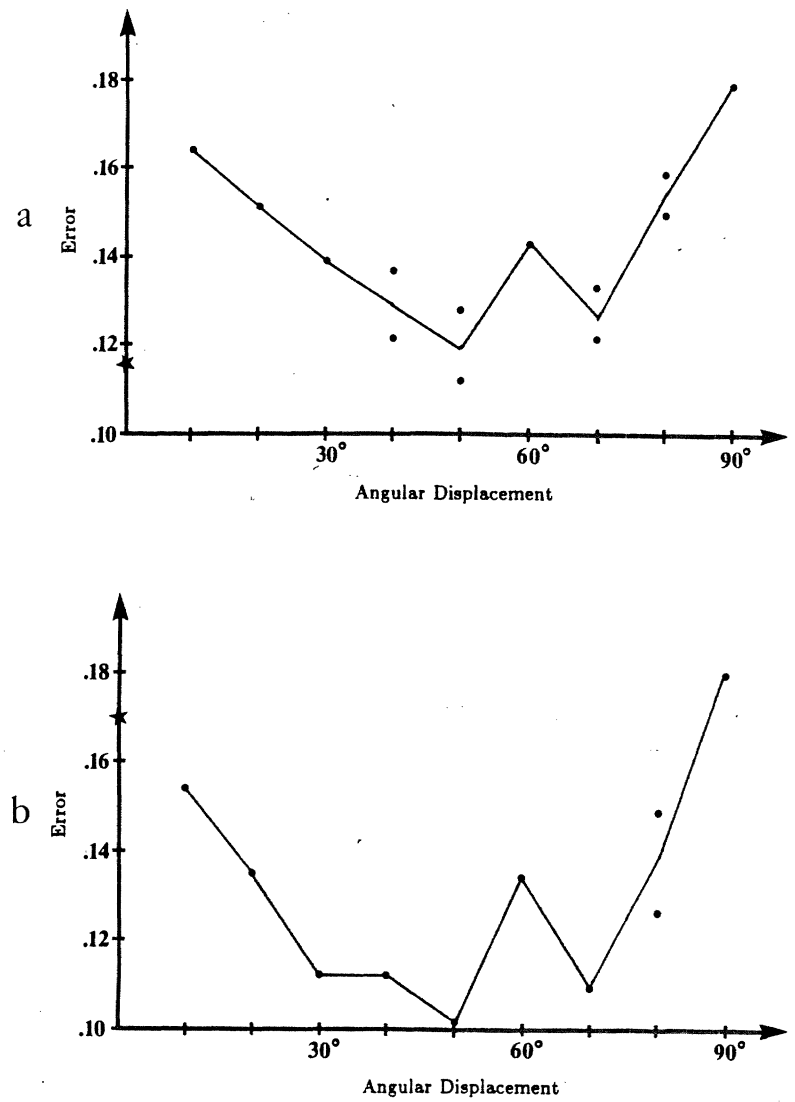

Figure 7. Absolute error in the internal distances between points in the 3-D model computed by the incremental rigidity scheme. The graph shows the error after a total of (a) 180° of rotation, and (b) 360° of rotation. The errors are again averaged over 10 random configurations of five points. The stars along the ordinate indicate the data points for the average errors of the continuous formulation after 180° and 360°.

To simplify the theoretical analysis, we consider in this section a measure of overall deviation from rigidity that is somewhat different from the one used by Ullman, and it is the following:

$$D_d(t,t') = \sum_{i,j} (l_{ij}^2(t) - l_{ij}^2(t'))^2. \tag{3.3}$$

This new measure uses the difference between the squares of the distances, $l_{ij}^2(t)$, rather

than the difference between the $l_{ij}(t)$'s themselves, thus avoiding the use of square roots. This measure also does not contain the cubic factor, $l_{ij}^3(t)$, in the denominator, which was included in the measure proposed by Ullman as a means of reducing the contributions of distant points relative to nearby ones. In this theoretical analysis, we mainly consider configurations of points that are compact, in the sense that the internal distances between pairs of points do not differ much from one another. In this case, the cubic factor seems not to be important.

The main problem that this theoretical analysis addresses is the long term stability of the solutions obtained by the incremental rigidity scheme, as a function of the angular separation between successive frames, for the case of rotation of rigid objects about a single axis in space. In the previous section, we observed the variation of convergence rate with angular displacement, in the results of computer simulations. In the present analysis, we examine this convergence phenomenon through a study of the stability of the incremental rigidity scheme. In particular, we analyze the behavior of the internal model under small perturbations, when it is near the true solution. We first examine the limiting case, where the frames are arbitrarily close to one another, that is when $t' \rightarrow t$. The analysis of this case is presented in sections 3.2.1 and 3.2.2. We then explore the convergence properties of the discrete formulation in sections 3.2.3 and 3.2.4. Finally, in section 3.2.5 we summarize the theoretical properties of the two schemes, as a function of angular displacement.

3.2.1 The Continuous Formulation

In the case of the continuous formulation, $l_{ij}(t')$ can be determined to a good first order approximation by $l_{ij}(t)$ and its derivative, $\dot{l}_{ij}(t)$, that is:

$$l_{ij}(t') \approx l_{ij}(t) + \dot{l}_{ij}(t) \times (t' - t). \tag{3.4}$$

As described earlier, when such an approximation is used, rather than computing the depth values $z_i(t')$ that maximize rigidity, we can compute the z components of velocity $\dot{z}_i$(t) that do the same task. It can be shown that in the limit, as $t' \rightarrow t$, using Eq. (3.3), and the definition of l_{ij}, the quantity to be minimized is:

$$D_c(t) = \sum_{i,j}((x_i - x_j)(\dot{x}_i - \dot{x}_j) + (y_i - y_j)(\dot{y}_i - \dot{y}_j) + (z_i - z_j)(\dot{z}_i - \dot{z}_j))^2. \tag{3.5}$$

Thus, given the model at time t, and the x and y components of velocity in the image, we compute the temporal derivatives of the depth values that minimize $D_c(t)$ given by Eq. (3.5). This is the continuous formulation that we use in the theoretical analysis of the convergence properties of the incremental rigidity scheme.

Note again, that the quantities that can be computed are not the absolute values of $\dot{z}_i(t)$ but the relative values $\dot{z}_i(t) - \dot{z}_j(t)$. It follows that without loss of generality, we can set (x_0, y_0, z_0) at the origin of the coordinate system, that is $(x_0, y_0, z_0) = (0, 0, 0)$. At every instant, the remaining $n - 1$ points are then given relative to this first point. Thus the number of independent variables in the model is $n - 1$. Introducing the following notational simplification:

$$a_{ij} = (x_i - x_j)(\dot{x}_i - \dot{x}_j) + (y_i - y_j)(\dot{y}_i - \dot{y}_j), \tag{3.6}$$

Eq. (3.5) can then be rewritten as follows:

$$D_c(t) = \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} (a_{ij} + (z_i - z_j)(\dot{z}_i - \dot{z}_j))^2 + \sum_{j=1}^{n-1} (a_{0j} + z_j \dot{z}_j)^2. \tag{3.7}$$

As described above, we are looking for the $\dot{z}_i(t)$ that minimize $D_c(t)$. The necessary condition for such a minimization is that the partial derivatives of $D_c(t)$ with respect to the $\dot{z}_i(t)$ are zero:

$$\frac{\partial D_c}{\partial \dot{z}_i(t)} = 0. \tag{3.8}$$

In section 3.5 we show that Eqs. (3.8) represent a set of $n-1$ linear equations with $n-1$ variables $\dot{z}_i(t)$, which generally have a unique solution. It follows that because $D_c(t) \geq 0$, thus being bounded from below, the $\dot{z}_i(t)$ that satisfy Eqs. (3.8) also minimize $D_c(t)$, so that these equations generally represent a sufficient condition for the minimization.

The $\dot{z}_i(t)$ that satisfy Eqs. (3.8) are expressed in terms of $z_i(t)$, thus representing a system of $n-1$ differential equations with $n-1$ variables. This system, however, is generally difficult to solve explicitly, as it is nonlinear. The only solutions that we were able to verify by straight substitution are the true motion of the object and its depth reversal (which are equivalent under orthographic projection). In the next section we present a stability analysis of the true rigid motion solution; that is, we examine the stability of the algorithm when its solution is close to the true solution. (The computer simulations address the full convergence behavior of the algorithm.) We concentrate on rigid motion, because we feel that nonrigid transformations may disrupt the stability of the solutions, introducing instabilities due to peculiar types of motion. For example, the internal model may be too sluggish in adjusting itself when fast changes of structure occur. By considering rigid motion in this analysis, we address more fundamental aspects of the theory itself.

3.2.2 Stability Analysis of the Continuous Formulation

The idea of a stability analysis can be stated as follows. Suppose that at a given instant t_0, the computed 3-D model is very close to the true solution, that is $z_i(t_0) = \hat{z}_i(t_0) + \epsilon_i(t_0)$ where $\hat{z}_i(t)$ is the true depth value at point i. Because the system is perturbed at t_0, we expect it in general to be perturbed for every $t \geq t_0$, that is $z_i(t) = \hat{z}_i(t) + \epsilon_i(t)$. The system is said to be asymptotically stable if the following is true:

$$\lim_{t \to \infty} \epsilon_i(t) = 0, \tag{3.9}$$

for every i. If the $\epsilon_i(t)$ remain bounded, but do not converge to zero, the system is defined to be weakly stable. If, however, $\lim_{t \to \infty} \epsilon_i(t) = \infty$ for some i, the system is said to be unstable.

If we suppose that $\epsilon_i(t)$ is small enough for every $t \geq t_0$, then we can make the following first order approximation of Eqs. (3.8) in terms of $\epsilon_i(t)$ and $\dot{\epsilon}_i(t)$ (the arguments, t, have been removed for simplicity):

$$\sum_{j=1}^{n-1} \left(\frac{\partial^2 D_c}{\partial \dot{z}_i \partial z_j} \epsilon_j + \frac{\partial^2 D_c}{\partial \dot{z}_i \partial \dot{z}_j} \dot{\epsilon}_j \right) = 0. \tag{3.10}$$

A derivation of Eqs. (3.10) is given in Appendix B. Eqs. (3.10) can be readily solved for $\dot{\epsilon}_j(t)$ in terms of $\epsilon_j(t)$. For notational simplicity, we write this solution in matrix form:

$$\dot{\underline{\epsilon}} = -\left[\frac{\partial^2 D_c}{\partial \dot{z}_i \partial \dot{z}_j} \right]^{-1} \left[\frac{\partial^2 D_c}{\partial \dot{z}_i \partial z_j} \right] \underline{\epsilon} = \mathbf{A}\underline{\epsilon}, \tag{3.11}$$

where $\underline{\epsilon} = (\epsilon_1, \ldots, \epsilon_{n-1})^T$, $[\partial^2 D_c / \partial \dot{z}_i \partial z_j]$ is the matrix whose element at row i and column j is $\partial^2 D_c / \partial \dot{z}_i \partial z_j$ and $[\partial^2 D_c / \partial \dot{z}_i \partial \dot{z}_j]^{-1}$ is the inverse of the matrix whose element at row i and column j is $\partial^2 D_c / \partial \dot{z}_i \partial \dot{z}_j$, provided it exists. The matrix $\mathbf{A}$ is defined to be $-[\partial^2 D_c / \partial \dot{z}_i \partial \dot{z}_j]^{-1} [\partial^2 D_c / \partial \dot{z}_i \partial z_j]$, and is evaluated at the true solution itself at every instant.

Using the Eqs. (3.11) to study the stability of the system, it is not possible to prove that the system is unstable. To do so requires a proof that $\underline{\epsilon}$ is unbounded as $t \to \infty$, but in this case the first order approximation used to derive Eqs. (3.10) no longer holds. It is still possible to determine, however, whether or not the system is asymptotically stable. Indeed, we will show in this section, that for many types of motion, the continuous formulation of the incremental rigidity scheme is not asymptotically stable. The results of computer simulations provide evidence that the formulation is, in general, not unstable, thus being weakly stable.

The Eqs. (3.10) represent a system of ordinary linear differential equations, which are used in conjunction with Eq. (3.7) where $D_c(t)$ is defined. Note therefore, that $\mathbf{A}$ has time dependent components. How do the different types of motion determine $\mathbf{A}$? A generalized motion of a rigid body can be described instantaneously as a rotation about a fixed axis in space plus a translation of the origin of the coordinates. We noted before that under orthographic projection, all that can be determined are the relative depths of the points and not their absolute values. Also, in the same case, the only relevant data to the problem of recovering structure from motion are the relative image coordinates (terms a_{ij} in Eq. (3.7)). Translation does not change any relative distances, as it changes the positions of all points by the same amount. Thus only rotations need to be considered in this problem.

As mentioned earlier, the matrix $\mathbf{A}$ in Eq. (3.11) is in general time dependent. Thus this system cannot be solved by the standard method of characteristic values, available to systems with constant coefficients. Furthermore, the general rotation can have variable angular velocity, and the axis of rotation can change over time, making Eq. (3.11) very difficult to integrate analytically.

It is not necessary, however, to solve Eq. (3.11) in order to reach certain conclusions regarding the nonconvergence of $\underline{\epsilon}$ to 0. For example, let $\underline{\epsilon}_j(t)$ be $n-1$ arbitrary solutions of Eq. (3.11) with initial conditions $\underline{\epsilon}_j(0)$ specified. Then the $n-1 \times n-1$ matrix $\mathbf{E}(t) =$

$[c_1(t), ..., c_{n-1}(t)]$ satisfies the Liouville Jacobi formula (Yakubovich and Starzhinskii, 1975):

$$Det(\mathbf{E}(t)) = Det(\mathbf{E}(0))exp\left(\int_0^t Tr(\mathbf{A}(t'))dt'\right). \tag{3.12}$$

It follows that a necessary condition for $\lim_{t \to \infty} \mathbf{E}(t) = 0$, that is, for the system to be asymptotically stable, is:

$$\int_0^\infty Tr(\mathbf{A}(t'))dt' = -\infty. \tag{3.13}$$

We found that condition 3.13 does not hold for many interesting types of motion. For example, for any three point configuration, rotating with arbitrary angular velocity about an axis perpendicular to the viewing axis, we can derive the following relationship, by using Eqs. (3.7) and 3.11 (details of the derivation are given in Appendix C):

$$Tr(\mathbf{A}(t)) = -\frac{(2\hat{z}_2 - \hat{z}_1)\dot{\hat{z}}_2 + (2\hat{z}_1 - \hat{z}_2)\dot{\hat{z}}_1}{\hat{z}_2^2 - \hat{z}_1\hat{z}_2 + \hat{z}_1^2}, \tag{3.14}$$

where $\hat{z}_1, \hat{z}_2, \dot{\hat{z}}_1, \dot{\hat{z}}_2$ denote the z coordinates and velocities evaluated at time t. This trace is periodic and has zero integral over the cycle, for rotations of fixed angular velocities ω:

$$\int_0^{2\pi/\omega} Tr(\mathbf{A}(t))dt = 0, \tag{3.15}$$

a result that contradicts Eq. (3.12). Thus, for any three point configuration have this type of motion, the internal 3 D model computed by the continuous formulation will not converge to the true structure.

The fact that this result is derived for three point structures is nontrivial. In fact, it is shown in the analysis of the discrete formulation, later in this section, that converging models can be constructed for three point configurations under certain conditions. A detailed analysis for structures with a higher number of points is in general very cumbersome, but we verified that Eq. (3.15) holds for special four point configurations, such as those that when viewed along the rotation axis, appear as squares, rectangles or trapezoids.

We also note that having the rotation axis perpendicular to the viewing axis is the best situation, as far as convergence is concerned, because depth motion is lost as the rotation axis is slanted towards the viewing axis (that is, away from the image plane). In particular, note that in the limiting case, where the rotation axis is parallel to the viewing axis, no motion in depth occurs whatsoever, and so the structure cannot be recovered from relative motion.

Is the instability of the continuous formulation due to properties of the projections of the object at particular positions in its revolution, which somehow convey less information to the recovery of structure from motion (for example, when many of the points belonging to an object have little motion in depth)? In order to check this hypothesis, we explored the convergence of the internal model when the object was oscillated back and forth

over a small angular extent, around different positions. It was found again, by checking the validity of Eq. (3.15), that regardless of position, the internal model is unable to converge to the true structure, thus giving a negative answer to the question raised. This suggested to us that the reason underlying the instability of the continuous formulation may be based on the smallness of the angular displacement between consecutive frames, and not on the structure of the frames themselves. Pursuing this idea further, the next sections address the convergence behavior of the discrete (and more general) formulation of the incremental rigidity scheme, focusing on the effects of the size of the angular displacements between frames.

3.2.3 The Discrete Formulation

The continuous formulation was derived from the approximation of Eqs. (3.3), as the angular displacements between the consecutive frames used in the structure from motion process become small, and it was expressed in Eq. (3.5). If we do not make such an approximation we have a more general formulation, which is however discrete. By being more general, this formulation allows a study of the incremental rigidity scheme as a function of the angular displacement between frames, and enables us to verify whether the total instability found in the continuous formulation is due to an "artifact" of the approximation made, namely an infinitesimal angular displacement between frames, or rather is due to a continuous process in which the system becomes less stable with smaller angles and eventually becomes unstable when the displacement is small enough.

In order to stress the discrete nature of the general formulation, we first rewrite Eq. (3.3), using as independent variable $t = k\tau$, where k is an integer and τ a fixed interval:

$$\begin{aligned} D_d(k\tau) = \sum_{i,j} \big[& (x_i(k\tau) - x_j(k\tau))^2 + (y_i(k\tau) - y_j(k\tau))^2 + (z_i(k\tau) - z_j(k\tau))^2 \\ & - (x_i((k+1)\tau) - x_j((k+1)\tau))^2 - (y_i((k+1)\tau) - y_j((k+1)\tau))^2 \\ & - (z_i((k+1)\tau) - z_j((k+1)\tau))^2 \big]^2 . \end{aligned} \tag{3.16}$$

Note that as in the continuous formulation, the quantities of interest for the theory are relative rather than absolute, i.e. $(z_i((k+1)\tau) - z_j((k+1)\tau))$. Thus again, without loss of generality we can set $(x_0, y_0, z_0) = (0, 0, 0)$, and study the behavior of the other $n-1$ points relative to it. For notational simplicity, let:

$$\begin{aligned} b_{ij}(k\tau) = & (x_i((k+1)\tau) - x_j((k+1)\tau))^2 + (y_i((k+1)\tau) - y_j((k+1)\tau))^2 \\ & - (x_i(k\tau) - x_j(k\tau))^2 - (y_i(k\tau) - y_j(k\tau))^2 . \end{aligned} \tag{3.17}$$

We now rewrite Eq. (3.16), using this notational simplification and without explicitly specifying τ:

$$D_d(k) = \sum_{i=1}^{n}\sum_{j=i+1}^{2n-1}(b_{ij}(k) + (z_i(k+1) - z_j(k+1))^2 - (z_i(k) - z_j(k))^2)^2 + \sum_{j=1}^{n-1}(b_{0j}(k) + z_j^2(k+1) - z_j^2(k))^2. \tag{3.18}$$

In the present formulation we are looking for the $z_i(k+1)$ that minimize $D_d(k)$. As in the continuous case, the necessary and sufficient condition for this minimization is:

$$\frac{\partial D_d}{\partial z_i(k+1)} = 0, \tag{3.19}$$

which represents a set of $n-1$ nonlinear equations implicitly relating the $z_i(k)$ and $z_i(k+1)$. The only instance in which we have been able to derive a full analytic solution for these equations is in the two point case. This case is a degenerate one, because for two points the 3 D structure is not determined uniquely by any number of views. It is important to discuss this case, however, as it provides a relevant insight into the theory, which is discussed further in section 3.2.5. Eqs. (3.18) and (3.19) in this case reduce to:

$$\left[b_{01}(k) + z_1^2(k+1) - z_1^2(k)\right] z_1(k+1) = 0. \tag{3.20}$$

The only nontrivial solution for this equation occurs when the term inside the brackets is zero. But using the definition of b_{01}, this condition essentially implies that the distances between the two points is constant. In other words, the distance between the points in the initial internal model at $k = 0$ will tend to remain the same throughout the entire motion even if it is wrong. The model will only correct itself if its initial guess for the distance between the points is small. In that case, the internal model will be forced to change because the distance between the points in the projected plane will become larger than the distance in the model. In this case, numerical calculations show that the model expands, and even has a small overshoot such that the distance between the points becomes a little larger than the true distance, and then stays at this condition forever.

3.2.4 Stability Analysis for the Discrete Formulation

The stability analysis for the discrete formulation begins with the same general argument as for the continuous case. Suppose that at $k = 0$, the model is very close to the true solution, i.e. $z_i(0) = \hat{z}_i(0) + \epsilon_i(0)$, where $\hat{z}_i(k)$ is the true depth value. We examine whether or not the perturbation at later times converges to zero, that is, whether $\lim_{k\to\infty} \epsilon_i(k) = 0$. If this is the case, and the $\epsilon_i(k)$ always remain small, then we can write an equation similar to Eq. (3.11) for the discrete formulation:

$$\underline{\epsilon}(k+1) = -\left[\frac{\partial^2 D_d}{\partial z_i(k+1)\partial z_j(k+1)}\right]^{-1}\left[\frac{\partial^2 D_d}{\partial z_i(k+1)\partial z_j(k)}\right]\underline{\epsilon}(k) = \mathbf{B}(k)\underline{\epsilon}(k). \tag{3.21}$$

Again, by the same arguments given in the continuous cases, we cannot use Eq. (3.21) to prove the system to be unstable, but only verify whether its convergence is asymptotic.

In this analysis we concentrate on three point configurations, because as in the continuous formulation, an analytic study with a larger number of points is too cumbersome. In the results of computer simulations, however, we found that the results derived here seem to generalize readily to configurations with a higher number of points. We found in all the cases studied, that if the viewed object is rotating with constant angular velocity around an axis perpendicular to the viewing axis, then the discrete formulation yields a converging internal model, but that the rate of convergence depends on the angular displacements between consecutively viewed frames.

In order to define the concept of rate of convergence, we first point out that recursive use of Eq. (3.21) shows that $\underline{\epsilon}(k)$ can be readily computed from $\underline{\epsilon}(0)$:

$$\underline{\epsilon}(k) = (\prod_{j=0}^{k-1} \mathbf{B}(j))\underline{\epsilon}(0), \tag{3.22}$$

and thus the convergence of $\underline{\epsilon}(k)$ to 0 depends only on this matrix multiplication, rather than on the perturbations themselves. Furthermore, if $\mathbf{B}(j)$ is cyclic (as it is for rotations in which the ratio between the angular displacement and 2π is rational), with cycle m, convergence depends only on the multiplications of the matrices over the cycle, which we denote:

$$\mathbf{B}_m = \prod_{j=0}^{m-1} \mathbf{B}(j). \tag{3.23}$$

Interestingly, the behavior of $\mathbf{B}_m^j$ bears a strong similarity with geometric sequences. Let us define the spectral radius of $\mathbf{B}_m$, $\rho(\mathbf{B}_m)$, as the maximal modulus of the characteristic values of $\mathbf{B}_m$. Then it is possible to show (see Appendix D) that:

$$\rho^j(\mathbf{B}_m) = \rho(\mathbf{B}_m^j). \tag{3.24}$$

From this result, one can understand the following important conclusion (see, for example, Varga, 1962):

$$\lim_{j \to \infty} \mathbf{B}_m^j = 0 \Leftrightarrow \rho(\mathbf{B}_m) < 1. \tag{3.25}$$

In other words the necessary and sufficient condition for the internal model to be asymptotically stable around the true solution, is that the spectral radius of the rotation matrix, $\mathbf{B}_m$, is less than 1.

Furthermore, from Eq. (3.24), one sees that the spectral radius corresponds to a "time constant" by which we can measure the velocity of convergence of the internal model. In fact, as ρ becomes closer to 1, more revolutions are necessary to obtain the same amount of convergence of the system, and the spectral radius of $\mathbf{B}_m^j$ declines exponentially with j. This exponential decline, however, holds only for the case of small perturbations, from which Eq. (3.22) was derived, and describes only a general trend, as in fine detail, the cyclic matrix $\mathbf{B}_m$ is composed of a multiplication of partial matrices

that may impose a "noise" in the decline. This noise can be seen in the simulations shown in section 3.1.

The dependence of the rate of convergence of the internal model on the angular displacement between viewed frames, d, and its consequences, are better illustrated by some examples. Fig. 8 displays ρ as a function of d for the case in which the viewed object appears as an equilateral triangle, when viewed along the rotation axis.

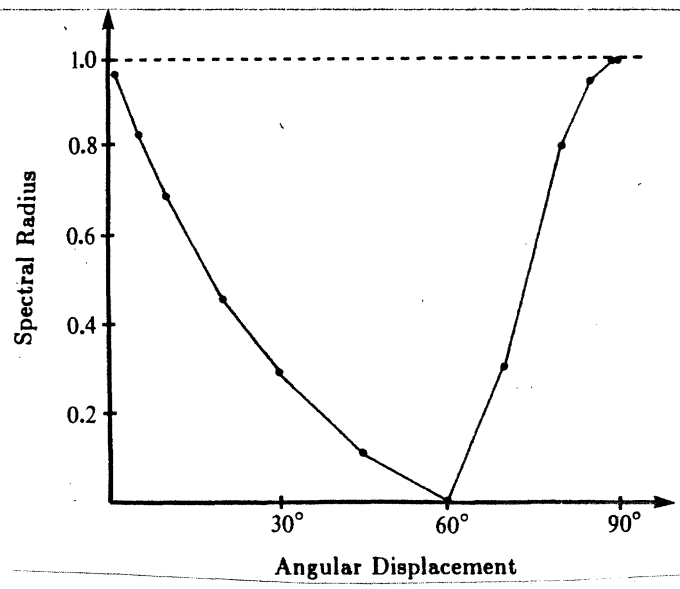

Figure 8. Graph of the spectral radius ρ as a function of d for the case in which the viewed object appears as an equilateral triangle, when viewed along the rotation axis.

The first result of interest in this graph is that $\rho < 1$ for every angular displacement d such that $0° < d < 90°$. This means that for almost every displacement used by the algorithm, the true solution (up to a depth reversal) is an asymptotically stable one. This result held for every configuration tried, provided that the rotation axis was perpendicular to the viewing axis. Thus, in spite of the fact that the true solution is asymptotically stable in the case of the discrete formulation, in the limit of arbitrarily closed frames, that is, in the continuous formulation, the true solution is not asymptotically stable. As can be seen in the figure, the spectral radius tends to 1 as $d \to 0$. Thus, in the limit, the condition expressed in Eq. (3.25) for the convergence of $\mathbf{B}^j_m$ to 0, is not fulfilled. This confirms our previous results regarding the lack of stability for the continuous formulation.

In addition, from the limit of 1 at small displacements, ρ decreases with d, showing a steady improvement in the rate of convergence of the internal model to the true solution with displacements up to 60°. This observation was also made by Ullman (1984). The behavior of ρ with small displacements is examined more closely in the next section. In the rest of this section we discuss the affect of large angular displacements on the spectral radius.

For large displacements, the rate of convergence deteriorates as can be seen in Fig. 8, approaching the value of 1 for $d = 90°$. The spectral radius is 1 for displacements

of 90° because only two different views of the object are available under orthographic projection, which is too little information for recovering structure from motion based on the assumption of rigidity. This result appears to be independent of other properties of the moving points.

This behavior of ρ with angular displacement can be related to the convergence rate observed in computer simulations. This is illustrated in Fig. 9. We began with the set of three points, whose true coordinates (when projected onto the $X - Z$ plane) lie around an equilateral triangle, as shown in Fig. 9a. The y coordinates of the points were nonzero but small. These points were then rotated around the vertical axis, and the error measure shown in Eq. (3.1) was computed, as a function of time. Fig. 9b shows an error plot for the case in which the angular displacement between frames was 20°, and the points were rotated for a total of 10 revolutions. The corresponding graph, with error plotted on a log scale, is shown in Fig. 9c. It can be seen that on a log scale, the long term convergence of the computed 3 D model is approximately linear. A measure of convergence rate can be given by the slope of the line that best fits this data, in a least squares sense. In Fig. 9d, the slope of this line is plotted as a function of the angular displacement between frames (the three point configuration was rotated for a total of 10 revolutions in each case). It can be seen that this graph is qualitatively similar to the graph of ρ shown in Fig. 8.

3.2.5 Convergence Under Small Angular Displacements and Instability of the Continuous Formulation

The question addressed in this section is, why does the rate of convergence fall with decreasing angular displacement between consecutive frames, eventually leading to instability for the case of the continuous formulation? One would certainly expect poorer corrections for the perturbations if smaller displacements are used, because less resolution is available in the image data. That is, the signal-to-noise ratio is substantially reduced. Many more views, however, are seen in the case of small angular displacements, which might trade off with the deterioration in the data resolution. One must keep in mind that there are two types of noise that can influence the performance of the incremental rigidity scheme: (1) the noise from the image data, and (2) the difference between the model and the true solution. In the computer simulations presented earlier, the image measurements were computed analytically, and hence had little error, so that the primary source of noise was the discrepancy between the computed model and true structure. This source of noise contributes to the deterioration of the solution with smaller angular displacements between frames. From the results in the previous section, we can conclude that in some sense, with decreasing displacement, the deterioration due to poorer corrections occurs at a faster rate than the improvement with the number of views. We explore this phenomenon in this section.

Formally, we can express the above question as, why does the spectral radius ρ of the rotation matrix $\mathbf{B}_m$ increase with decreasing angular displacement d, eventually tending to 1 when $d \rightarrow 0$? This is certainly a characteristic of the curve of ρ vs. d curve if d is small enough (see Fig. 8). The matrix $\mathbf{B}_m$ results from the multiplication

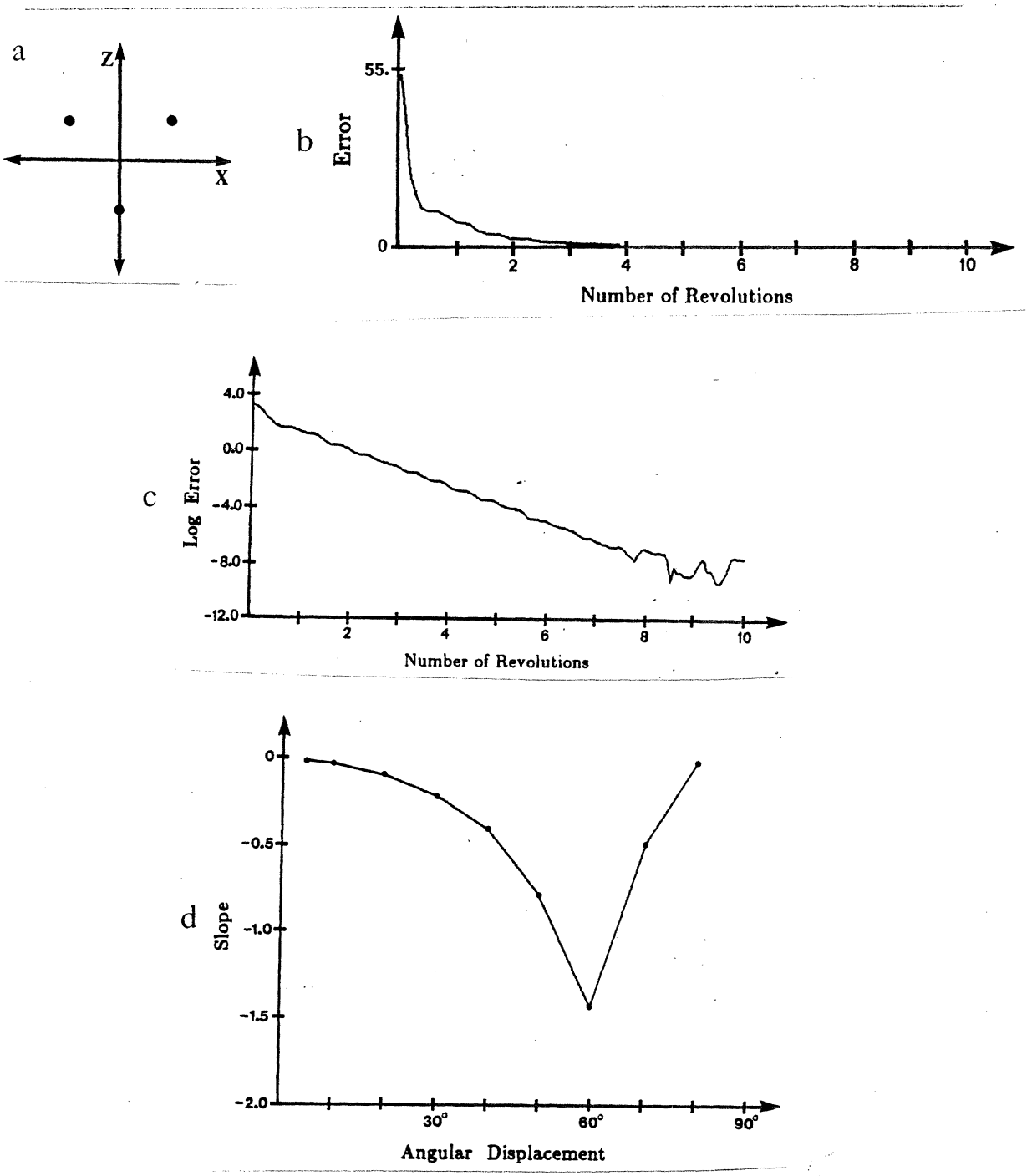

Figure 9. The results of computer simulations. (a) The true three point configuration, projected onto the $X - Z$ plane. (b) The error in the computed 3 D model, as a function of time, for 20° displacements between frames. (c) The error in the 3 D model, plotted on a log scale. (d) The rate of convergence of the solution, as a function of the angular displacement between frames.

of the matrices corresponding to partial displacements (Eq. (3.23)). One would like to define some quantity related to these partial displacement matrices, which expresses the deterioration of their information content, and to study the connection of this quantity to the general spectral radius as a function of d. This quantity and its relationship to the spectral radius were found by numerical experimentation. The intuition underlying these observations is what follows.

Given three point configurations and fixed angular displacements, an inspection of the partial displacement matrices shows that they all have exactly the same two characteristic values, c_1 and c_2, which are real, and $0 < c_1 < 1$ and $1 < c_2$. If these characteristic values belonged to a diagonal matrix, then the effect of such a matrix would be to change the perturbation vector $\underline{\epsilon} = (\epsilon_1, \epsilon_2)^T$ into $(c_1\epsilon_1, c_2\epsilon_2)$. In general, however, the vector $\underline{\epsilon}$ can point in any arbitrary direction, and the effects of a diagonal matrix on the length of $\underline{\epsilon}$ would be different from direction to direction. Thus a meaningful quantity belonging to these matrices must be some average of the effect of the matrix over all directions of the perturbation vector. An important example is the mean length squared of the vector resulting from the application of the matrix to the general unitary vectors, $(cos\gamma, sin\gamma)$. This quantity is the following:

$$\frac{1}{2\pi}\int_0^{2\pi}(c_1^2 cos^2(\gamma) + c_2^2 sin^2(\gamma))d\gamma = \frac{c_1^2 + c_2^2}{2}. \tag{3.26}$$

The partial displacement matrices are in general not diagonal, but if d is small enough they can be closely approximated as orthogonally similar to the diagonal matrix having c_1 and c_2 as its elements. If we begin with an arbitrary perturbation vector, these matrices will rotate the vectors and transform the resultants as described above, and then repeat this transformation until a cycle is completed. Thus for small enough displacements, one can expect the diagonal matrices to work on perturbation vectors of almost all directions, thus having an average effect as described in the last paragraph. Generally this mean effect always represents a decrease in the perturbation vector, which is applied $2\pi/d$ times during a cycle of revolution of the object. Thus for small enough displacements one can expect a connection between c_1, c_2, ρ and d given by an equation of the following form:

$$\rho(d) = \left(\frac{c_1^2(d) + c_2^2(d)}{2}\right)^{2\pi/d}. \tag{3.27}$$

Eq. (3.27) closely matches the behavior of the spectral radius, for the equilateral triangle case (Fig. 8) for $d < 20°$. For other configurations, relationships very similar to this one were observed to apply. The importance of Eq. (3.27) is that it relates a quantity that can be computed in any partial displacement matrix, namely $(c_1^2(d) + c_2^2(d))/2$ (the deterioration factor), to the spectral radius of the full rotation matrix, and it enables us to reduce the problem of why the continuous formulation is unstable to the question of how fast does the deterioration factor tend to 1 as $d \to 0$. In fact one can see from Eq. (3.27) that a necessary condition for $\lim_{d \to 0}\rho(d) = 1$ is that:

$$\lim_{d \to 0} \frac{c_1^2(d) + c_2^2(d)}{2} = 1. \tag{3.28}$$

This condition, however, is not sufficient. For example, if the deterioration factor approaches 1 linearly with d, i.e. $(c_1^2(d) + c_2^2(d))/2 \approx 1 - \alpha d$ for small d, then because of the power $2\pi/d$ in Eq. (3.27) we would have $\lim_{d \to 0} \rho(d) = e^{-2\pi\alpha} < 1$. We expect then, that the deterioration factor approaches 1 faster than linear. This is indeed the case, as can be seen by expanding $(c_1^2(d) + c_2^2(d))/2$ into the first few powers of its Taylor series. For any three point configuration, this expansion yields:

$$\frac{c_1^2(d) + c_2^2(d)}{2} \approx 1 - \beta d^2. \tag{3.29}$$

This quadratic approach of the deterioration to 1 is fast enough to ensure that $\lim_{d \to 0} = 1$, as can be seen by substituting Eq. (3.29) into Eq. (3.27) and taking the limit.

Therefore, we confirm that the instability of the continuous formulation is due to the fact that the deterioration in the data resolving power as a result of the small angular displacements between frames is faster (quadratic with the inverse of the displacement) than the increase of information due to the increase in the number of frames (linear with the inverse of the displacement). This analysis provides theoretical support for the observation that for the case of the discrete formulation of the incremental rigidity scheme, the rate of convergence of the internal model to the true solution diminishes as the angular displacements between frames decreases. In the limit of very small displacements, that is, in the case of the continuous formulation, this deterioration leads to a lack of asymptotic convergence of the model to the true structure.

4. Perspective Formulations of the Incremental Rigidity Scheme

In this section, we first present discrete and continuous formulations of the incremental rigidity scheme that use perspective projection of the scene onto the image plane. We then present a few examples of the results of computer simulations, in which the perspective formulations were applied to sequences of points undergoing both pure rotation about a single axis in space and pure translation through space. The behavior of the perspective formulation of the incremental rigidity scheme is more complex than that of the orthographic formulation. We found that if the absolute position of an object in space is known throughout the motion of the object, then the perspective formulation performs well, similar to the orthographic formulation. If the absolute position of the object is unknown, the incremental rigidity scheme in general does not derive the correct 3-D structure. The results of computer simulations revealed a degradation in performance with smaller angular and spatial displacements between frames, but this degradation was somewhat more severe than in the orthographic formulation. In the limit of the continuous formulation, the solution is no longer stable.

In both the discrete and continuous formulations, we assume that the positive z axis points in the direction of the optical axis, with the image plane at $z = 1$, as shown in Fig. 10. For notational convenience, we let $(\hat{x}_i(t), \hat{y}_i(t), \hat{z}_i(t))$ represent the true coordinates

of point i in space and $(u_i(t), v_i(t))$ denote the true projected image coordinates of point i. The image coordinates are given by:

$$u_i(t) = \frac{\hat{x}_i(t)}{\hat{z}_i(t)} \qquad v_i(t) = \frac{\hat{y}_i(t)}{\hat{z}_i(t)}. \tag{4.1}$$

The coordinates of the points in the computed model are denoted by $(x_i(t), y_i(t), z_i(t))$.

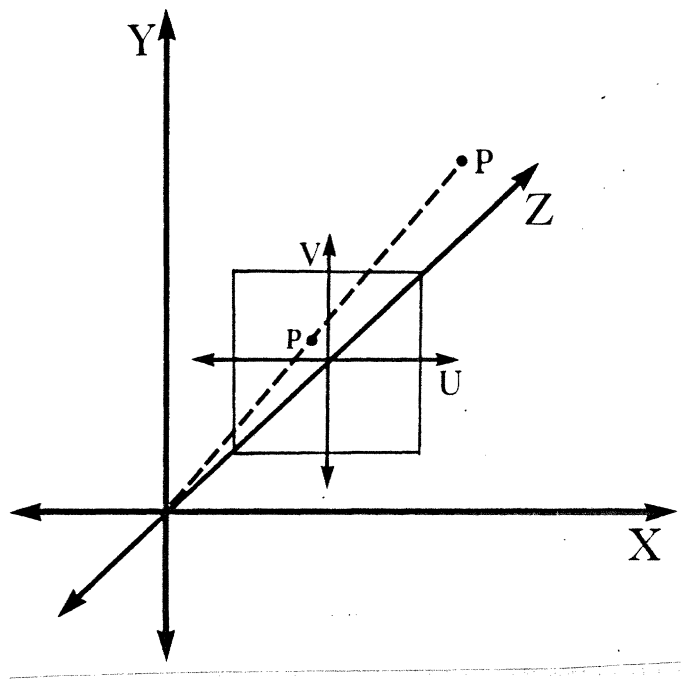

Figure 10. The imaging geometry for perspective projection. The positive z axis points in the direction of the optical axis, with the image plane at $z = 1$.

4.1 The Discrete Formulation

In the discrete case, we compute the depth values $z_i(t')$ that minimize the measure of rigidity, $D_d(t, t')$ given by the following (for notational simplicity, (x_i, y_i, z_i) refer to the coordinates of the points at time t in the computed model, while (x'_i, y'_i, z'_i) refer to the coordinates of the points at a later time t'):

$$\begin{aligned} D_d(t,t') &= \sum_{i,j} (l_{ij}(t) - l_{ij}(t'))^2 \\ &= \sum_{i,j} \Big[\left((x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2\right)^{\frac{1}{2}} \\ &\quad - \left((x'_i - x'_j)^2 + (y'_i - y'_j)^2 + (z'_i - z'_j)^2\right)^{\frac{1}{2}} \Big]^2 . \end{aligned} \tag{4.2}$$

This measure is the same as that used in the orthographic formulation (i.e., it is derived from Eq. (2.1) with the $l_{ij}^3(t)$ factor in the denomonator omitted, as shown in Eq. (A1) of Appendix A).

As before, we assume that we have a current model $(x_i(t), y_i(t), z_i(t))$. Rather than assuming that $x_i(t')$ and $y_i(t')$ are known, however, we assume that only $u_i(t')$ and $v_i(t')$ are known (they are derived from the true space coordinates at time t'). The coordinates of the points in space at time t' for the computed model are then given implicitly by:

$$\begin{aligned} x_i(t') &= u_i(t')z_i(t') \\ y_i(t') &= v_i(t')z_i(t'). \end{aligned} \tag{4.3}$$

By substituting Eqs. (4.3) into Eq. (4.2), $D_d(t,t')$ can be expressed in terms of the coordinates of the points as follows ((u_i', v_i') denote the image coordinates at time t'):

$$\begin{aligned} D_d(t,t') = \sum_{i,j} \Big[&\left((x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2\right)^{\frac{1}{2}} \\ &- \left((u_i'z_i' - u_j'z_j')^2 + (v_i'z_i' - v_j'z_j')^2 + (z_i' - z_j')^2\right)^{\frac{1}{2}} \Big]^2 . \end{aligned} \tag{4.4}$$

After the values $z_i(t')$ that minimize $D_d(t,t')$ are computed, Eqs. (4.3) are used to derive the space coordinates $x_i(t')$ and $y_i(t')$. The new 3 D coordinates, $(x_i(t'), y_i(t'), z_i(t'))$, then become the current 3 D model, a new frame in the sequence is registered, and the process repeats itself.

In the case of perspective projection, 3 D structure can be recovered from relative motion only up to a multiplicative scale factor. This is clear by inspection of Eqs. (4.1); a constant scaling of the space coordinates $(x_i(t), y_i(t), z_i(t))$ does not change the image coordinates $(u_i(t), v_i(t))$. In most of the computer simulations, we assumed that the overall scale factor is known; this is equivalent to assuming that the absolute position of the object in space is known. In the simulations, the absolute coordinates of one of the n points in motion was supplied to the algorithm and the coordinates of the remaining $n-1$ points were computed relative to the known point (similar to the simulations of the orthographic formulation). If the absolute position of the object in space is not known, then the scale of the computed 3 D model depends on the choice of the initial configuration. For example, suppose that we assume that the set of moving points initially lies on a plane that is parallel to the image plane. The scale of the computed 3 D model will then depend on where this initial plane is placed in depth. That is, if the plane is positioned twice as far from the image plane, the computed 3 D model will essentially be twice as large. We will briefly mention the results of computer simulations in which the absolute position of the object in space is unknown. In these simulations, when we began with a flat initial configuration, we usually placed the initial z coordinates of the points at a depth that was approximately the mean of the true z coordinates of the points in space. Further details of the implementation of the discrete formulation can be found in Appendix A.

4.2 The Continuous Formulation

In the continuous case, we compute the z components of velocity that minimize the measure of rigidity, $D_c(t)$, given by the following (the coordinates and velocities are all measured at time t):

$$D_c(t) = \sum_{i,j} \frac{[(x_i - x_j)(\dot{x}_i - \dot{x}_j) + (y_i - y_j)(\dot{y}_i - \dot{y}_j) + (z_i - z_j)(\dot{z}_i - \dot{z}_j)]^2}{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}. \tag{4.5}$$

This measure is the same as that used in the orthographic case (see Eq. (2.7)). It is assumed that a current model of the coordinates of the points in space is given, along with the known coordinates and velocities of the points in the image plane, $u_i(t), v_i(t), \dot{u}_i(t), \dot{v}_i(t)$. The velocities of the points in space for the computed model are then given implicitly by:

$$\begin{aligned} \dot{x}_i(t) &= u_i(t)\dot{z}_i(t) + \dot{u}_i(t)z_i(t) \\ \dot{y}_i(t) &= v_i(t)\dot{z}_i(t) + \dot{v}_i(t)z_i(t). \end{aligned} \tag{4.6}$$

If we let $x_{ij} = x_i - x_j$, $y_{ij} = y_i - y_j$, $z_{ij} = z_i - z_j$, and $\dot{z}_{ij} = \dot{z}_i - \dot{z}_j$, then $D_c(t)$ is expressed in terms of the coordinates of the points as follows:

$$D_c(t) = \sum_{i,j} \frac{[x_{ij}(u_i\dot{z}_i + \dot{u}_i z_i - u_j\dot{z}_j - \dot{u}_j z_j) + y_{ij}(v_i\dot{z}_i + \dot{v}_i z_i - v_j\dot{z}_j - \dot{v}_j z_j) + z_{ij}\dot{z}_{ij}]^2}{x_{ij}^2 + y_{ij}^2 + z_{ij}^2}. \tag{4.7}$$

In other respects, the continuous formulation is similar to the discrete formulation. A model of the structure of the moving object is built up by continually taking into account new velocity information over an extended time period. After the $\dot{z}_i(t)$ are computed by minimizing Eq. (4.7), Eqs. (4.6) are used to derive $\dot{x}_i(t)$ and $\dot{y}_i(t)$, which are then used to compute the new 3-D model. Again, because perspective projection is used, the structure of the object can only be computed up to a multiplicative scale factor. Further details of the implementation of the continuous formulation are presented in Appendix A.

4.3 Computer Simulations

In this section, we present some results of the computer simulations of the discrete and continuous perspective formulations of the incremental rigidity scheme, for a small number of examples. We first consider the case in which a set of discrete points is translated through space, and then consider the rotation of a set of points about a vertical axis. Through simulations of the discrete formulation, we examine the rate of convergence of the algorithm and the quality of the solution that it yields, as a function of the size of the angular and spatial displacements between frames. We then observe the limiting behavior of the continuous formulation for the case of rotation. We also note the difference in performance of the perspective formulations when the absolute position of the moving points in space is either known or unknown.

It should be stressed that in the case of orthographic projection, it is not possible to recover 3 D structure for objects undergoing pure translation, because the relative positions of projected points in the image do not change as objects translate. The use of perspective projection has the advantage of allowing the recovery of structure for translating objects. In the simulations presented here, objects were oscillated back and forth in the direction parallel to the image plane. If the objects were allowed to translate in one direction for a large extent, the projected image of the points would eventually become so small that the recovery of structure would become difficult due to a loss of numerical accuracy in the measurement of the changing positions of the points. Oscillating the points allows us to analyze their structure over an extended time period while maintaining a relatively large image.

In the first set of examples, the input consisted of a set of six points in space. The coordinates of the points were chosen randomly and the coordinates of one of the points was known to the algorithm. The projections of the points onto the $X - Y$ and $X - Z$ planes are shown in Figs. 11a and 11b, respectively. This set of points was oscillated back and forth in a direction parallel to the image plane. The $X - Z$ projection (birds' eye view) of the rightmost and leftmost position of the points in space are shown in Figs. 11c and 11d, respectively. The x coordinates of the points in Figs. 11c and 11d are shifted by 60 units from the initial positions shown in Fig. 11b. In the simulations, the initial points shown in Fig. 11b were first translated to the right in constant discrete steps and then translated to the left with the same discrete steps. A 3 D model was built up over several oscillations of the points. The initial model for the structure of the points was flat, with the z coordinates placed near the mean of the true z coordinates of the points. For the particular set of points used in these examples, $z_i(t) = 80$, $i = 1, ..., n$, in the initial 3 D model. Once the initial $z_i(t)$ are specified, the initial $x_i(t)$ and $y_i(t)$ can be determined from the image coordinates $u_i(t)$ and $v_i(t)$. The one point whose position is known throughout the oscillation of the points is indicated by an arrow in Fig. 12a.

Fig. 12 shows a birds' eye view of the computed 3-D model after 1 (left column), 4 (middle column) and 10 (right column) complete oscillations of the points, using three different sizes of spatial displacements between frames. The true 3-D structure of the points is shown again in Fig. 12a. In the case of Fig. 12b, the size of the spatial displacements between frames was 60, so that each complete oscillation consisted of four frames (i.e., the points were displaced to the right in one step, returned to the central position, displaced to the left in one step and then returned again to the central position). In the case of Fig. 12c, the points were translated in steps of 30, so that a complete oscillation of the points consisted of 8 frames. Finally, in the case of Fig. 12d, the points were translated in steps of 5, so that each full oscillation consisted of 48 frames. In Fig. 12e, we show plots of the error in the computed models as a function of time, for 10 complete oscillations of the points. The error measure used here is the same as that used in the orthographic case, shown in Eq. (3.2). The error plots for spatial displacements of 60, 30 and 5 are shown superimposed, with the displacements indicated above each curve. From the results shown in Fig. 12, it can first be seen that the algorithm does essentially converge to the true structure of the points. Second, the rate of convergence of the computed model to the true solution decreases with smaller spatial displacements.

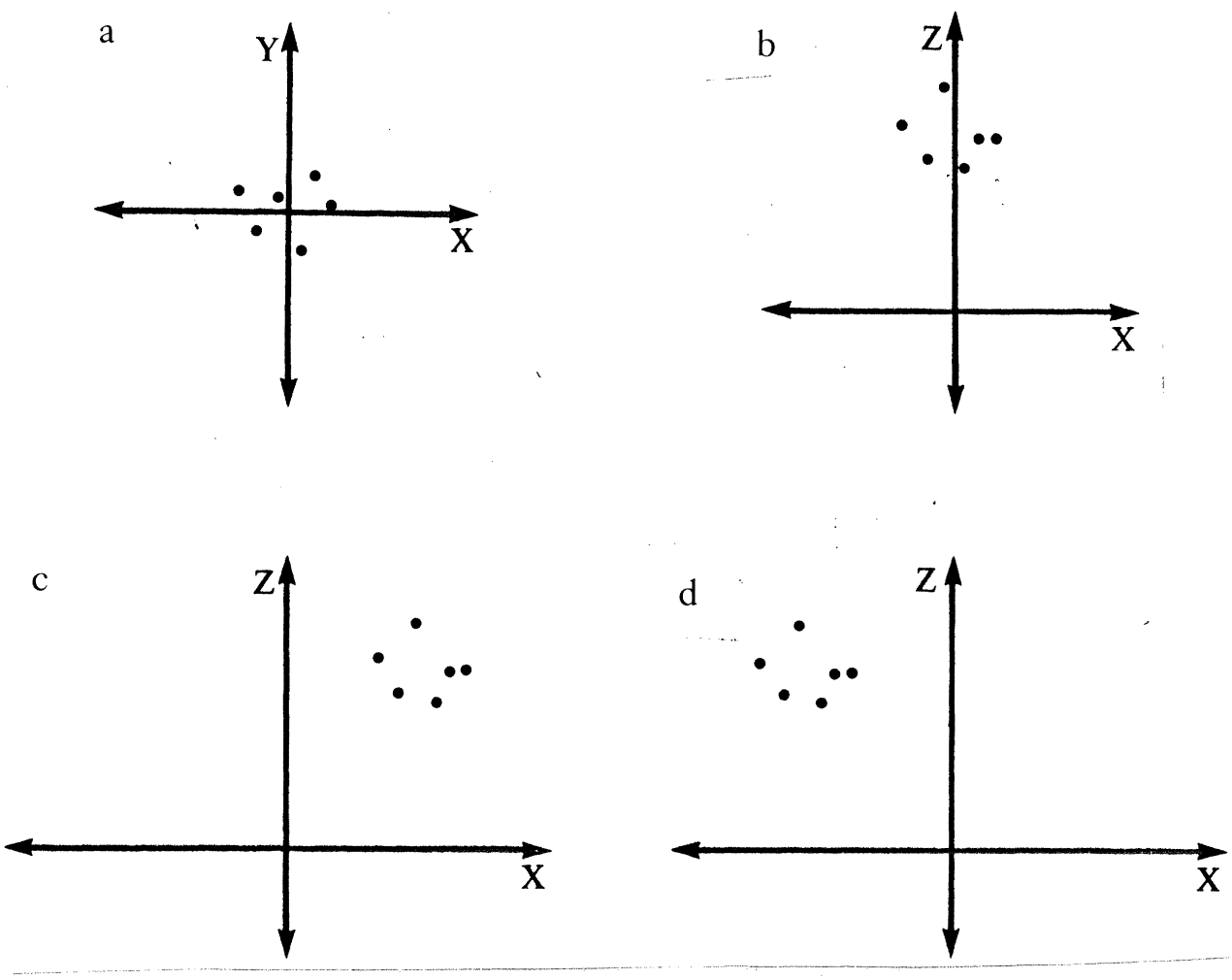

Figure 11. A set of six points with random coordinates is shown projected onto (a) the $X - Y$ plane and (b) the $X - Z$ plane. (c) The rightmost position of the points during their oscillations back and forth in the direction parallel to the image plane. (d) The leftmost position of the points.

For example, after a single oscillation (leftmost column), the model shown in Fig. 12b (displacements of 60) is clearly closer to the true structure of the points shown in Fig. 12a than the model shown in Fig. 12d (displacements of 5). The three error plots shown in Fig. 12e also have periodic variations superimposed on a steady decline in error, with the amplitude of the variations increasing with smaller spatial displacements. Thus, in the case of the perspective formulation, it appears that the performance of the incremental rigidity scheme degrades with smaller spatial displacements when objects undergo pure translation through space. This result is analogous to the observation that in the case of

orthographic projection, the performance of the scheme degrades with smaller angular displacements when objects undergo rotation.

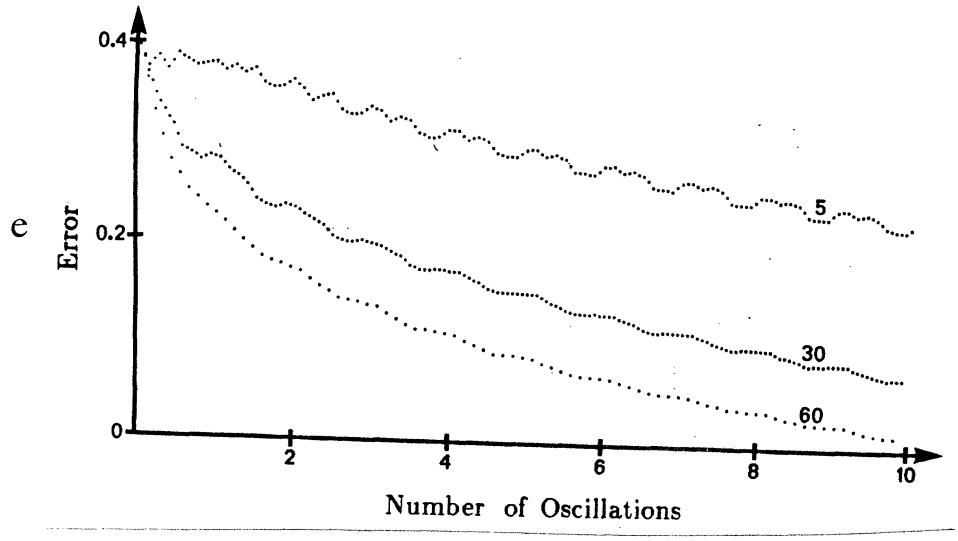

Figure 12. (e) Graphs of the error in the distances between points in the computed 3 D model, as a function of time. The points underwent 10 complete oscillations, in steps of 60, 30 and 5.

Fig. 13 illustrates the behavior of the perspective formulation, for the case of rotation of the points about a central vertical axis. The set of six points shown in Fig. 11 was rotated about a seventh point that was placed at the position:

$$(x_i(t), y_i(t), z_i(t)) = (0, 0, 80).$$

The axis of rotation passed through this point and was perpendicular to the $X - Z$ plane. The position of this central point was known to the algorithm throughout the motion of the points and the positions of the remaining six points were computed relative to this central point. As before, the initial configuration was flat, with $z_i(t) = 80$, $i = 1, ..., n$.

The set of 7 points was rotated in angular steps of 40° and 5°. Figs. 13b and 13c show the computed 3 D model after total rotations of 40°, 80°, 120° and 360° (indicated on the left), for angular displacements between frames of 40° and 5°, respectively. The true positions of the points are shown in Fig. 13a. In Fig. 13d, the graphs of the error in the computed models are shown as a function of total angular rotation, for three full revolutions of the points. It can be seen that the rate of convergence and quality of the 3 D structure decreases with the smaller angular displacement between frames. Although similar to the case of orthographic projection, the degradation in performance with smaller angles was somewhat more severe in the case of perspective projection.

Fig. 14 illustrates the behavior of the continuous formulation of the incremental rigidity scheme, for the case of perspective projection. The same set of points shown in Fig. 11 was rotated about the vertical axis and the 3 D model was computed at infinitesimally closely spaced frames, using the instantaneous velocities projected onto

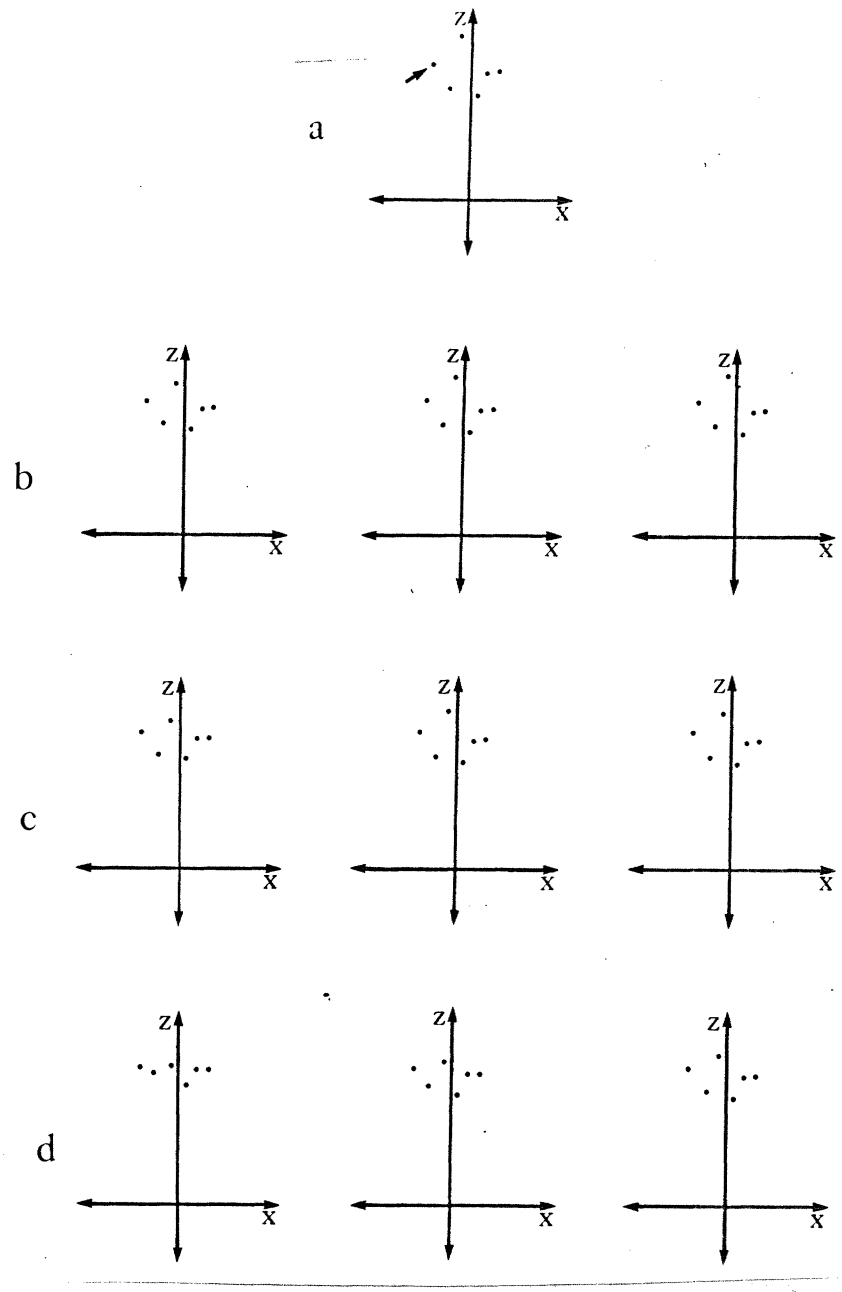

Figure 12. (a) Projection onto the $X - Z$ plane of the true 3-D structure of the set of points. (b) The computed 3-D model after 1 (left column), 4 (middle column) and 10 (right column) complete oscillations of the points, for a spatial displacement between frames of size 60. (c) and (d) The same three computed 3-D models, for spatial displacements of size 30 and 5, respectively.

the image. As in the previous rotation examples, a seventh point was placed at the center of rotation of the points, and the position of this central point was known to the algorithm throughout the rotation of the points. In Fig. 14a, we compare the true positions of the points (left) with the best solution (right) obtained over four full revolutions of the points. Although the model is quite close to the true structure at this position of the points, the solution oscillates significantly over an extended time period, similar to the orthographic case. A graph of the error in the computed model over the four revolutions is shown in Fig. 14b. The arrow marks the point at which the solution shown in Fig. 14a was obtained. There is an initial fast convergence toward the true structure of the points, but the algorithm then oscillates with high amplitude away from the true solution. We conclude that similar to the orthographic case, the direct use of velocity information for the recovery of structure by the incremental rigidity scheme does not lead to a robust solution.

We should finally note that we also explored some examples in which no information about the absolute position of the points in space is known. In this case, we began with an initial configuration that was flat, and otherwise provided no further constraint on the positions of the points. The points were either oscillated back and forth in the direction parallel to the image plane, or rotated about a vertical axis. We found that the incremental rigidity scheme would sometimes derive the correct solution under these conditions, but in general, the algorithm did not derive the correct structure, although the computed solution was essentially rigid. The computed solution was not just a scaled or rotated version of the true structure, but actually a different one altogether. This suggests that some additional constraint may be required to derive the correct solution.

To summarize, the perspective formulation of the incremental rigidity scheme appears to perform well when the absolute position of the object in space is known. This formulation then allows the recovery of structure for objects undergoing pure translation through space, as well as rotation. The results of computer simulations revealed a degradation in performance with smaller angular and spatial displacements between frames. This degradation appeared to be more severe than in the case of orthographic projection, when the points were rotated about a vertical axis. When the points were translated, there was a more gradual decrease in the rate of convergence and quality of the computed 3-D model with smaller spatial displacements. In the limit of the continuous formulation, the solution was no longer stable.

5. Summary and Conclusions

In this paper, we studied and generalized the incremental rigidity scheme for the recovery of structure from motion. This algorithm, as first proposed by Ullman (1984), assumed the visual input to consist of a sequence of frames, each containing a finite number of points. The scheme maintained an internal model of the structure of the viewed object, which was updated from frame to frame, so as to be consistent with the changing image and to be as rigid as possible. This internal model was shown to correctly converge to the true structure of the object, for both rigid and nonrigid motions.

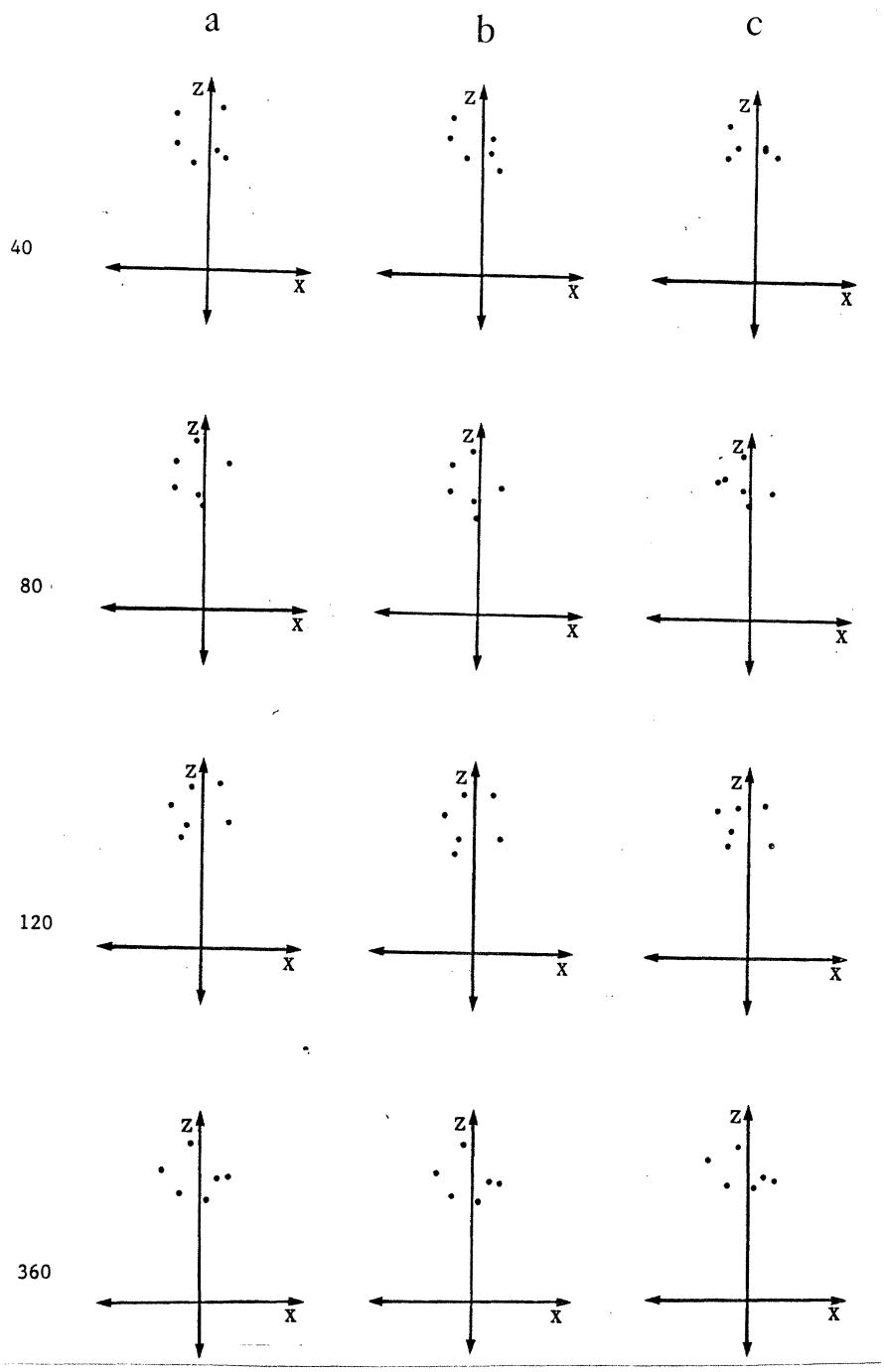

Figure 13. (a) The true positions of the points after rotations by 40°, 80°, 120° and 360°. (b) The computed 3 D model after rotation of the points by 40°, 80°, 120° and 360°, for an angular displacement between frames of 40°. (c) The same computed 3 D models, for an angular displacement of size 5°. ((d) on next page.)

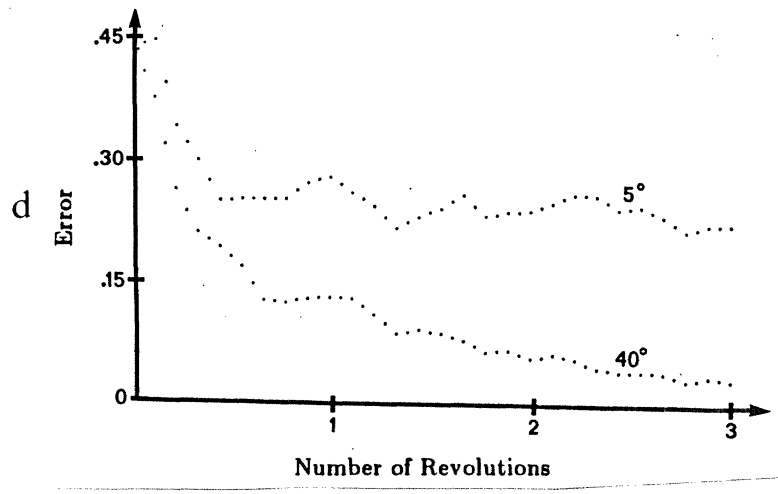

Figure 13. (d) Graphs of the error in the distances between points in the computed 3 D model, as a function of time, for three full revolutions of the points.

In the present paper, we focused on an observation made by Ullman, that the rate of convergence of the internal model to the true structure increased with the amount of change between consecutive frames. A major part of our analysis focused on objects that rotated with constant angular velocity around an axis perpendicular to the viewing axis, and whose image was formed through an orthographic projection, at equally spaced angular displacements. The complete dependence of the convergence rate on the size of the angular change was obtained for a large variety of objects.

Ullman's observation was confirmed for small displacements. Indeed, the convergence rate first increased towards a maximum and then decreased, as a function of the angular displacements between the frames. The deterioration of the algorithm for high displacements can be understood if we note that the number of frames available for analysis in a given total number of revolutions of the object, or in other words, the amount of information provided to the scheme, is reduced in inverse proportion to the angular changes between the frames.

The deterioration of the algorithm for small displacements, however, has a more complex and surprizing basis. In this case, although the number of frames used in a given amount of rotation is high, the spatial resolution between the points is low (see discussion on the beggining of section 3.2.5). Our analysis indicates that the latter effect dominates for very small angular jumps between frames, reducing the convergence rate. Furthermore, we found that this reduction is such that in the limit of infinitesimally small displacements, where only velocity information is used, the internal model does not even have a full convergence to the true structure of the viewed object, but only a rough approach to the correct solution.

In analogy to linear algebra, one can say that the transformations from frame to frame of the internal model in the incremental rigidity scheme become ill conditioned as

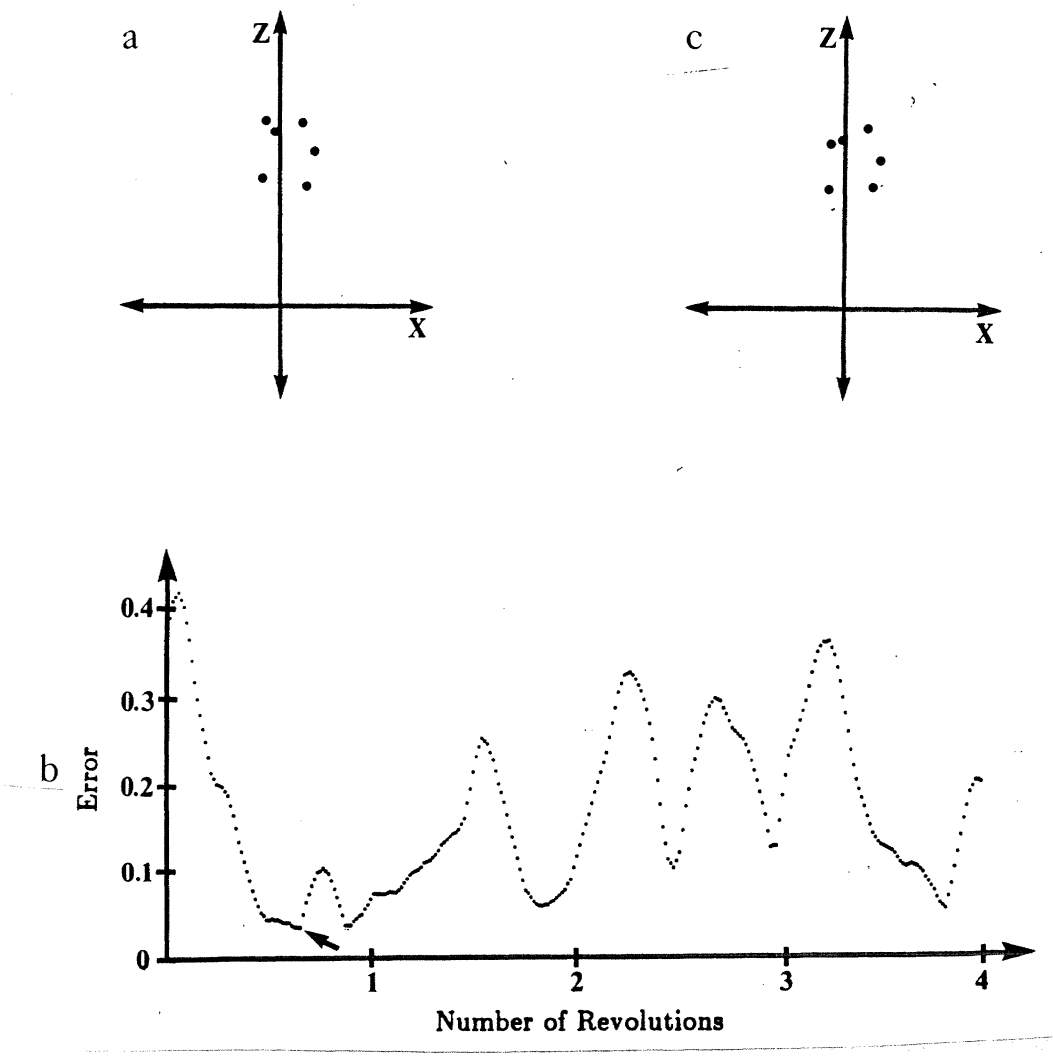

Figure 14. (a) The best solution (right) obtained over four revolutions of the points is compared with the true positions (left), for the case of the continuous formulation. (b) The error in the internal 3-D distances between points in the model, as a function of time. The arrow in (b) indicates the instant at which the solution in (a) was obtained.

the angular displacements decrease, i.e. they become sensitive to noise. This property is not unique to the present transformations, but occur in other important situations such as numerical differentiation. The noise sensitivity of a differentiation of order m increases as n^m when the number of points n in a given interval is increased (Strang, 1976).

As mentioned before, the orthographic projection is in general not physically valid. It was used in the present paper, because it allows a deeper analysis of the phenomena studied. This analysis is further validated because these results are not limited to the

case of orthographic projection, or to the case in which the object rotates. Under certain conditions, the incremental rigidity scheme can be generalized successfully to formulations that use perspective projection, and that in this case, the structure of the viewed object is recoverable both for rotation and translation. Also, similar to the orthographic case, the convergence rate of the internal model to the true structure decreased, if smaller spatial changes between consecutive frames were used.

Inspection of previous results by other researchers suggests that the use of information that is local in space and time is insufficient to allow a robust recovery of structure from motion. We distinguish between two types of local information that are relevant to the problem: (1) spatial locality, which refers to the use of a small number of points or features of the viewed object, and (2) temporal locality, which is the use of a small number of views of the object. Early algorithms for recovering structure from motion based this recovery on a limited number of points and views, and were able to recover the structure of the viewed object when it was rigid. When an object deformed, some algorithms could determine its nonrigidity through inconsistencies of the underlying equations. It followed that these schemes are not robust against noise, as noise has similar effects as nonrigidity.

It can be shown that motion information that is extended in space, but temporally local, can be used to provide stable recovery of structure from motion (Bruss and Horn, 1983; Negahdaripour and Horn, 1985). Perceptual evidence, however, suggests that spatial extension is not necessary for the solution of the structure from motion problem by the human visual system (Borjesson and von Hofsten, 1973; Johansson, 1975; Petersik, 1980).

On the contrary, Ullman's incremental rigidity scheme, which was applied locally in space, both in its original formulation and in the present paper, uses an internal model that is updated constantly, in order to extend motion information over time, and converge to the correct solution. Some perceptual studies suggest that this temporal extension may be used by the human visual system (Wallach and O'Connell, 1953; White and Mueser, 1960; Green, 1961).

We showed in the present paper, however, that an algorithm that overcomes temporal locality does not necessarily provide a robust solution to the structure from motion problem. It seems that in this case, a necessary condition for a stable solution is the use of views of the object that are significantly disparate. In the limit, the use of pure velocity information allows only a rough solution to the structure from motion problem, which is less stable over an extended time period. A somewhat similar conclusion was derived by Ullman (1983), who argued that in a temporally local scheme, the recovery of structure from the instantaneous velocity field is impossible under orthographic projection, and that for perspective projection, the recovery is unstable. We suggest that the problems with using local velocity information are too large to be overcome by an extension of this information over time. In other words, the inaccuracies introduced by a velocity based scheme cannot be corrected by the use of multiple views of the object.

Our results therefore hint that even in a temporally extended algorithm, a memory of sufficiently distinct past views or internal models of the viewed object is necessary for a robust recovery of its structure. We point out that the incremental rigidity scheme, in

its original formulation and in the present paper, used memory of only one past internal model. Although in principle, a memory of many past models could be used by the algorithm, it is remarkable that such a minimal amount of past information is sufficient to provide a robust recovery of structure. Still, it may be the case that the use of multiple past views or internal models for establishing a new model of 3 D structure could increase the rate of convergence of the internal model.

In further support of the above conclusions, we note that in order to recover 3 D structure from measured image velocities, it is essential that these velocities at least approximate the true projected velocities of moving elements in space. In general, however, it is difficult to derive real projected velocities from the optical flow pattern on the eye or camera. Factors such as changing illumination, specularities, shadows, and rotation of an object surface relative to a light source, can create patterns of optical flow that do not correspond to the real movement of features on a physical surface. The difficulty of computing correct projected velocities provides an additional motivation for not basing the recovery of structure from motion directly on velocity information.

Our analysis raises a number of issues regarding the recovery of structure from motion in the human visual system. First, it is not clear whether the visual system achieves a stable solution to the structure from motion problem, or a rough solution such as that provided by a velocity based scheme. The more robust solution may not be essential if we consider that other perceptual cues, such as binocular stereo, may help to improve the quality of the 3 D solution obtainable by the structure from motion process. Further psycophysical experiments are required to examine this question.

A second point of interest is that our analysis suggests that if the visual system incorporates a robust solution to the structure from motion problem, it must be able to match corresponding points in very disparate views of moving objects. The displacements between corresponding points may be larger than the spatial limits proposed for the short-range perceptual process found for 2 D motion (Braddick, 1973, 1974, 1980; Pantle and Picciano, 1976; Petersik, Hicks and Pantle, 1978; Petersik and Pantle, 1979). A similar conclusion, based on different grounds, was formulated by Petersik (1980). He explored the sensations elicited by stroboscopic simulations of a rotating transparent sphere filled with randomly positioned dots. By manipulating both spatial and temporal variables in the simulation, he concluded that corresponding elements in consecutive frames can be matched over spatial and temporal distances that exceed the empirically determined limits of the 2 D short range process. Even in this experiment, however, points that reach the periphery of the sphere have small displacements from frame to frame that may fall spatially inside the short-range process. It remains, therefore, to be established that these points do not provide all the information used by the subjects to sense the continuous rotation and internal volume of the sphere.

If the human visual system is able to match very distant corresponding points from disparate views of a moving object, the question of how we solve this correspondence problem becomes important (for a review of the motion correspondence problem, see Ullman (1981)). This correspondence could be established either through the repeated use of a short range matching process, or through the use of an explicit long-range matching process. In the first case, the short range process could provide an essentially

continuous tracking of features in the changing image, and the positions of features could then be sampled by a longer range tracking process that feeds disparate positions into the structure from motion process (a similar idea has been suggested by Ullman (1981)). In the latter case, a correspondence would be established directly from disparate views of moving objects, without the aid of the intermediate short range process. Petersik's (1980) results may suggest that it is possible to solve this long range correspondence problem for recovering structure. This correspondence problem could be solved in conjunction with the structure from motion process; that is, a correspondence could be chosen that subsequently gives rise to the most rigid 3 D interpretation (such a matching process requires a global decision procedure, and may be well suited to solution by parallel schemes such as Hopfield's "neuronal nets" (1984)). Given the difficulty of solving this long range correspondence problem in general, however, it seems unlikely that short range measurements would not be used when they are available.

Acknowledgements: We thank Tomaso Poggio, Shimon Ullman and Victor Inada for valuable comments on this paper, and Carol Bonomo, who kindly donated to us a megabucks ticket, and so maintained our dreams high.

References

E. H. Adelson, "Rigid objects that appear highly non rigid," Invest. Ophthalmol. Vision Sci. Suppl. **26**, 56, 1985.

D. Ballard and C. Brown, *Computer Vision* (Prentice Hall, New Jersey, 1982).

B. M. Bennett and D. D. Hoffman, "The computation of structure from fixed axis motion: rigid structures," unpublished manuscript, 1984a.

B. M. Bennett and D. D. Hoffman, "The computation of structure from nonrigid motion," unpublished manuscript, 1984b.

A. Bobick, "A hybrid approach to structure from motion," Proc. ACM Interdisciplinary Workshop on Motion: Representation and Perception, Toronto, Canada, 91 109, 1983.

E. Borjesson and C. von Hofsten, "Visual perception of motion in depth: application of a vector model to three dot motion patterns," Percept. Psychophys. **13**, 169 179, 1973.

O. Braddick, "The masking of apparent motion in random dots patterns," Vision Res. **13**, 355 369, 1973.

O. Braddick, "A short range process in apparent motion," Vision Res. **14**, 519 527, 1974.

O. Braddick, "Low level and high level processes in apparent motion," Phil. Trans. Royal Soc. London **290**, 137 151, 1980.

M. L. Braunstein, "Depth perception in rotation dot patterns: effects of numerosity and perspective," J. Exp. Psychol. **6**, 41 420, 1962.

M. L. Braunstein, *Depth Perception Through Motion* (Academic Press, New York, 1976).

A. Bruss and B. K. P. Horn, "Passive navigation," Comput. Vision Graph. Image Proc. **21**, 3 20, 1983.

W. F. Clocksin, "Perception of surface slant and edge labels from optical flow: a computational approach," Percept. **9**, 253 269, 1980.

J. J. Gibson and E. J. Gibson, "Continuous perceptive transformations and the perception of rigid motion," J. Exp. Psychol. **54**, 129 138, 1957.

B. F. Green, "Figure coherence in the kinetic depth effect," J. Exp. Psychol. **62**, 272 282, 1961.

C. J. Hay, "Optical motions and space perception – an extension of Gibson's analysis," Psychol. Rev. **73**, 550 565, 1966.

E. C. Hildreth, *The Measurement of Visual Motion* (MIT Press, Cambridge, 1984a).

E. C. Hildreth, "The computation of the velocity field," Proc. Royal Soc. London B. **221**, 189 220, 1984b.

D. D. Hoffman and B. E. Flinchbaugh, "The interpretation of biological motion," Biol. Cybern. **42**, 195 204, 1982.

J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two state neurons," Proc. Nat. Acad. Sci. (U.S.A.) **81**, 3088 3092, 1984.

B. K. P. Horn, *Robot Vision* (MIT Press and McGraw Hill, Cambridge, 1985).

B. K. P. Horn and B. G. Schunck, "Determining optical flow," Artif. Intell. **17**, 185 203, 1981.

G. Jansson and G. Johansson, "Visual perception of bending motion," Percept. **2**, 321–326, 1973.

G. Johansson, "Perception of motion and changing form," Scand. J. Psychol. **5**, 181 208, 1964.

G. Johansson, "Visual motion perception," Sci. Am. **232**, 76 88, 1975.

G. Johansson, "Spatial constancy and motion in visual perception," in *Stability and Constancy in Visual Perception*, ed. W. Epstein (Wiley, New York, 1977).

G. Johansson, "Visual event perception," in *Handbook of Sensory Physiology*, eds. R. Held, H. W. Leibowitz and H. L. Teuber (Springer Verlag, Berlin, 1978).

J. J. Koenderink, "Shape from motion and bending deformations," J. Opt. Soc. Am. A **1**, 1265 1266, 1984.

J. J. Koenderink and A. J. Van Doorn, "Depth and shape from differential perspective in the presence of bending deformations," unpublished manuscript, 1984.

H. C. Longuet Higgins, "A computer algorithm for reconstructing a scene from two projections," Nature **293**, 133–135, 1981.

H. C. Longuet Higgins, "The role of the vertical dimension in stereoscopic vision," Percept. **11**, 377–386, 1983.

H. C. Longuet Higgins, "Visual ambiguity of a moving plane," J. Opt. Soc. Am. A **1**, 1215, 1984.

H. C. Longuet Higgins and K. Prazdny, "The interpretation of moving retinal images," Proc. Royal Soc. London B. **208**, 385–397, 1981.

D. G. Luenberger, *Introduction to Linear and Nonlinear Programming* (Addison Wesley, Reading, MA, 1973).

D. Marr, *Vision* (W. H. Freeman Co., San Francisco, 1982).

W. R. Miles, "Movement interpretations of the silhouette of a revolving fan," Am. J. Psychol. **43**, 392–505, 1931.

H.-H. Nagel, "Recent advances in image sequence analysis," Proc. Premier Colloque Image, Biarritz, 1984.

S. Negahdaripour and B. K. P. Horn, "Direct passive navigation," MIT Artificial Intelligence Laboratory Memo 821, 1985.

A. J. Pantle and L. Picciano, "A multistable movement display: evidence for two separate systems in human vision," Science **193**, 500 502, 1976.

J. T. Petersik, "The effect of spatial and temporal factors on the perception of stroboscopic rotation simulations," Percept. **9**, 271 283, 1980.

J. T. Petersik, K. I. Hicks, and A. Pantle, "Apparent movement of successively generated subjective figures," Percept. **7**, 371–383, 1978.

J. T. Petersik and A. Pantle, "Factors controlling the competing sensations produced by a bistable stroboscopic motion display," Vision Res. **19**, 143–154, 1979.

N. Sugie and H. Inagaki, "A computational aspect of kinetic depth effect," Biol. Cybern. **50**, 431–436, 1984.

G. Sperling, M. Pavel, Y. Cohen, M. S. Landy, and B. J. Schwartz, "Image processing in perception and cognition," in *Physical and Biological Processing of Images* eds. O. J. Braddick and A. C. Sleigh (Springer Verlag, Berlin, 1983).

G. Strang, *Linear Algebra and its Aplications* (Academic Press, New York, 1976).

J. T. Todd, "Visual information about rigid and nonrigid motion: A geometric analysis," J. Exp. Psychol. **8**, 238–252, 1982.

J. T. Todd, "The perception of three dimensional structure from rigid and nonrigid motion," Percept. Psychophys. **36**, 97–103, 1984.

R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of three dimensional motion parameters of rigid objects with curved surfaces," University of Illinois at Urbana-Champaign, Coordinated Science Laboratory Report R-921, 1981.

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

S. Ullman "Analysis of visual motion in biological and computer systems," IEEE Computer, August, 57–69, 1981.

S. Ullman, "Recent computational studies in the interpretation of structure and motion," in *Human and Machine Vision*, eds. J. Beck, B. Hope and A. Rosenfeld (Academic Press, New York, 1983).

S. Ullman, "Maximizing rigidity: the incremental recovery of 3-D structure from rigid and rubbery motion," Percept. **13**, 255–274, 1984.

S. Ullman, "The optical flow of planar surfaces," unpublished manuscript, 1985.

R. S. Varga, *Matrix Iterative Analysis* (Prentice-Hall, New Jersey, 1962).

H. Wallach and D. N. O'Connell, "The kinetic depth effect," J. Exp. Psychol. **45**, 205–217, 1953.

H. Wallach, A. Weisz, and P. A. Adams, "Circles and derived figures in rotation," Am. J. Psychol. **69**, 48–59, 1956.

A. M. Waxman and S. Ullman, "Surface structure and 3-D motion from image flow: a kinematic analysis," Internat. J. Robotics Res., in press, 1985.

J. A. Webb and J. K. Aggarwal, "Visually interpreting the motions of objects in space," Computer **14**, 40–49, 1981.

B. W. White and G. E. Mueser, "Accuracy in reconstructing the arrangement of elements generating kinetic depth displays," J. Exp. Psychol. **60**, 1–11, 1960.

V. A. Yakubovich and V. M. Starzhinskii, *Linear Differential Equations with Periodic Coefficients* (John Wiley & Sons, New York, 1975).

Appendices

Appendix A

In this appendix, we present some of the details of the implementation of the incremental rigidity scheme used to derive the simulation results presented in sections 3 and 4.

Orthographic Projection: The Discrete Formulation

Ullman formulated the incremental rigidity scheme as the computation of depth values $z_i(t')$ that minimize the measure of rigidity given by $D(t,t')$, as shown in Eq. (2.2). The analyses presented in this paper mainly consider rigid objects in motion, which are compact in the sense that the internal distances between pairs of points do not differ much from one another. In this case, the additional distance factor in the denomonator of the measure used by Ullman has little influence on the resulting solution obtained by the algorithm. We therefore removed this factor in most of our simulations, and minimized instead the following expression (the distances $l_{ij}(t)$ and $l_{ij}(t')$ are expressed in terms of the coordinates of the points, and (x_i, y_i, z_i) refer to the model coordinates at time t, while (x_i', y_i', z_i') refer to the model coordinates at time t'):

$$D_d(t,t') = \sum_{i,j} \Big[\left((x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2\right)^{\frac{1}{2}} - \left((x_i' - x_j')^2 + (y_i' - y_j')^2 + (z_i' - z_j')^2\right)^{\frac{1}{2}} \Big]^2 . \tag{A1}$$

The $z_i(t')$ that minimize $D_d(t,t')$ satisfy a system of nonlinear equations given by:

$$\frac{\partial D_d}{\partial z_i(t')} = 0, \qquad i = 1, \ldots, n-1. \tag{A2}$$

Rather than solving this system of equations explicitly, we solved them implicitly through the use of a steepest descent minimization algorithm for Eq. (A1), based on the gradient of $D_d(t,t')$ (Luenberger, 1973). The components of the gradient are given by the following:

$$\frac{\partial D_d}{\partial z_i(t')} = -2 \sum_j \frac{l_{ij}(t) - l_{ij}(t')}{l_{ij}(t')} (z_i'(t) - z_j'(t)). \tag{A3}$$

In the case of orthographic projection, only relative depths can be recovered. In the implementation, we placed the initial point $(x_0(t), y_0(t), z_0(t))$ at the origin of the coordinate system; that is, $(x_0(t), y_0(t), z_0(t)) = (0,0,0)$. At every instant, the remaining $n-1$ points were then given relative to this initial point. Unless otherwise stated, the initial configuration of points was flat, that is, $z_i(0) = 0$, for $1, \ldots, n-1$. From this configuration, the algorithm can move toward two equally likely configurations, one being the mirror image reflection of the other about the image plane. In other words, this

configuration is located at a saddlepoint in the solution space for the problem (the solution space is an n dimensional space in which a value for $D_d(t,t')$ is assigned to every possible combination of depth values $z(t')$), at which the gradient of $D_d(t,t')$ is zero. To make this gradient nonzero, the initial solution was perturbed slightly, thus causing the algorithm to move in a particular direction.

The steepest descent minimization method from nonlinear programming (see, for example, Luenberger, 1973) was used to compute the $z_i(t')$ for each new frame. The current depth values $z_i(t)$ were used as the initial solution for $z_i(t')$ to begin the minimization for the next frame. The objective function is given by $D_d(t,t')$, and its gradient by Eqs. (A3). Golden section search was used to perform the one dimensional minimization within the steepest descent method.

Orthographic Projection: The Continuous Formulation

In section 2, we presented a continuous formulation, which requires the computation of z components of velocity, $\dot{z}_i(t)$ that minimize the measure of rigidity given by $D_c(t)$, as expressed in Eq. (2.3). In the computer simulations, we used the particular measure of overall deviation from rigidity given by:

$$D_c(t) = \sum_{i,j} \frac{[(x_i - x_j)(\dot{x}_i - \dot{x}_j) + (y_i - y_j)(\dot{y}_i - \dot{y}_j) + (z_i - z_j)(\dot{z}_i - \dot{z}_j)]^2}{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}. \qquad (A4)$$

The $\dot{z}_i(t)$ that minimize $D_d(t)$ satisfy a system of linear equations given by:

$$\frac{\partial D_c}{\partial \dot{z}_i(t)} = 0, \qquad i = 1, \ldots, n-1. \qquad (A5)$$

These equations are of the form:

$$\frac{\partial D_c}{\partial \dot{z}_i} = 2 \sum_j \frac{(x_i - x_j)(\dot{x}_i - \dot{x}_j) + (y_i - y_j)(\dot{y}_i - \dot{y}_j) + (z_i - z_j)(\dot{z}_i - \dot{z}_j)}{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}(z_i - z_j) = 0. \qquad (A6)$$

In the case of orthographic projection, only relative z components of velocity can be recovered. As in the discrete case, we again placed the initial point $(x_0(t), y_0(t), z_0(t))$ at the origin of the coordinate system throughout the entire motion, so that $(x_0(t), y_0(t), z_0(t)) = (\dot{x}_0(t), \dot{y}_0(t), \dot{z}_0(t)) = (0,0,0)$. At every instant the remaining $n-1$ z components of velocity were then given relative to this initial point. Unless otherwise stated, we again began with a flat initial configuration in which $z_i(0) = 0$, for $1, \ldots, n-1$, which was perturbed slightly so that the gradient of $D_c(t)$ is initially nonzero.

At each moment, $\dot{z}_i(t)$ were obtained by solving a system of linear equations, for which we used the simple Gauss–Seidel relaxation method. The initial condition for the relaxation was usually the set of velocities computed in the previous iteration and $\dot{z}_i = 0$ for the first iteration. To integrate motion information over an extended time period, we then made use of the following approximations using $\dot{x}_i(t)$, $\dot{y}_i(t)$, and $\dot{z}_i(t)$:

$$
\begin{aligned}
x_i(t) + \dot{x}_i(t)\delta t &= x_i(t+\delta t) \\
y_i(t) + \dot{y}_i(t)\delta t &= y_i(t+\delta t) \\
z_i(t) + \dot{z}_i(t)\delta t &= z_i(t+\delta t).
\end{aligned} \tag{A6}
$$

Once the $\dot{z}_i(t)$ were computed, by minimizing $D_e(t)$, then the $z_i(t+\delta t)$ could be derived from the current model $z_i(t)$. The new model was then taken as $M(t+\delta t) = (x_i(t+\delta t), y_i(t+\delta t), z_i(t+\delta t))$, and the process was continued. The time interval, δt, typically corresponded to an angular displacement between frames on the order of 0.1 degrees. We also experimented with angular displacements up to two orders of magnitude smaller, and found the qualitative behavior of the algorithm to be the same.

Perspective Projection: The Discrete Formulation

In the case of perspective projection, the x and y coordinates at time t' can be expressed in terms of the known image coordinates at time t' and the depth values to be computed, $z_i(t')$, as follows:

$$
\begin{aligned}
x_i(t') &= u_i(t')z_i(t') \\
y_i(t') &= v_i(t')z_i(t').
\end{aligned} \tag{A7}
$$

The measure of rigidity to be minimized is given by substituting Eqs. (A7) into Eq. (A1):

$$
\begin{aligned}
D_d(t,t') = \sum_{i,j} \Big[&\left((x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2\right)^{\frac{1}{2}} \\
&- \left((u_i'z_i' - u_j'z_j')^2 + (v_i'z_i' - v_j'z_j')^2 + (z_i' - z_j')^2\right)^{\frac{1}{2}} \Big]^2 .
\end{aligned} \tag{A8}
$$

The $z_i(t')$ that minimize $D_d(t,t')$ satisfy a system of nonlinear equations given by:

$$
\frac{\partial D_d}{\partial z_i(t')} = 0, \qquad i = 1, \ldots, n-1. \tag{A9}
$$

In the computer simulations, these equations were solved implicitly through the use of a steepest descent minimization algorithm for Eq. (A8), based on the gradient of $D_d(t,t')$. The components of the gradient are given by the following:

$$
\frac{\partial D_d}{\partial z_i(t')} = -2 \sum_j \frac{l_{ij} - l_{ij}'}{l_{ij}'} \left[u_i'(u_i'z_i' - u_j'z_j') + v_i'(v_i'z_i' - v_j'z_j') + (z_i' - z_j')\right]. \tag{A10}
$$

The computed $z_i(t')$ are used to derive the new $x_i(t')$ and $y_i(t')$ using Eqs. (A7), which then provide the new computed 3 D model. This new model also serves as an initial solution to begin the minimization for the next frame.

Perspective Projection: The Continuous Formulation

In the continuous case, the x and y components of velocity of the points in space can be expressed in terms of the known image coordinates and velocities as follows:

$$\begin{aligned} \dot{x}_i(t) &= u_i(t)\dot{z}_i(t) + \dot{u}_i(t)z_i(t) \\ \dot{y}_i(t) &= v_i(t)\dot{z}_i(t) + \dot{v}_i(t)z_i(t). \end{aligned} \tag{A11}$$

If we let $x_{ij} = x_i - x_j$, $y_{ij} = y_i - y_j$, $z_{ij} = z_i - z_j$, and $\dot{z}_{ij} = \dot{z}_i - \dot{z}_j$, then $D_c(t)$ is expressed in terms of the coordinates of the points as follows:

$$D_c(t) = \sum_{i,j} \frac{[x_{ij}(u_i\dot{z}_i + \dot{u}_i z_i - u_j\dot{z}_j - \dot{u}_j z_j) + y_{ij}(v_i\dot{z}_i + \dot{v}_i z_i - v_j\dot{z}_j - \dot{v}_j z_j) + z_{ij}\dot{z}_{ij}]^2}{x_{ij}^2 + y_{ij}^2 + z_{ij}^2}. \tag{A12}$$

The $\dot{z}_i(t)$ that minimize $D_c(t)$ satisfy a system of linear equations given by:

$$\frac{\partial D_c}{\partial \dot{z}_i} = 0 \qquad i = 1, \ldots, n-1. \tag{A13}$$

These equations are of the form:

$$\frac{\partial D_c}{\partial \dot{z}_i} = 2\sum_j b_{ij}\,[u_i(x_i - x_j) + v_i(y_i - y_j) + (z_i - z_j)] = 0, \tag{A14}$$

where the b_{ij} are given by:

$$b_{ij} = \frac{x_{ij}(u_i\dot{z}_i + \dot{u}_i z_i - u_j\dot{z}_j - \dot{u}_j z_j) + y_{ij}(v_i\dot{z}_i + \dot{v}_i z_i - v_j\dot{z}_j - \dot{v}_j z_j) + z_{ij}\dot{z}_{ij}}{x_{ij}^2 + y_{ij}^2 + z_{ij}^2}. \tag{A15}$$

At each moment, the $\dot{z}_i(t)$ were obtained by solving the system of linear equations given by Eqs. (A14), for which we used the simple Gauss–Seidel relaxation method. After the $\dot{z}_i(t)$ were known, the Eqs. (A11) were used to derive the x and y components of velocity in space, $\dot{x}_i(t)$ and $\dot{y}_i(t)$. To integrate motion information over an extended time period, we then made use of the approximations given in Eqs. (A6) to compute a new model $(x_i(t+\delta t), y_i(t+\delta t), z_i(t+\delta t))$.

Appendix B

In this appendix, we explain how Eq. (3.10) was derived.

Eq 3.8 depends on the image coordinates (data) and on depth coordinates of the points $z_i(t)$ and their velocities $\dot{z}_i(t)$ (computed values). These depth variables are in general displaced from the true motion values. This departure can be incorporated in Eq. (3.8) as follows:

$$\frac{\partial D}{\partial \dot{z}_i}(t, \hat{z}_1 + \epsilon_1, \ldots, \hat{z}_{n-1} + \epsilon_{n-1}, \dot{\hat{z}}_1 + \dot{\epsilon}_1, \ldots, \dot{\hat{z}}_{n-1} + \dot{\epsilon}_{n-1}) = 0. \tag{B1}$$

If the c and $\dot{c}$ are small enough, as it is assumed in the stability analysis of section 3.2.2, then the first elements of the Taylor expansion of Eq. (B1) yields the following approximation:

$$\begin{aligned} 0 = & \frac{\partial D}{\partial \dot{z}_i}(t, \hat{z}_1, ..., \hat{z}_{n-1}, \dot{\hat{z}}_1, ..., \dot{\hat{z}}_{n-1}) \\ & + \sum_{j=1}^{n-1} \frac{\partial D}{\partial \dot{z}_i \partial z_j}(t, \hat{z}_1, ..., \hat{z}_{n-1}, \dot{\hat{z}}_1, ..., \dot{\hat{z}}_{n-1}) c_j \\ & + \sum_{j=1}^{n-1} \frac{\partial D}{\partial \dot{z}_i \partial \dot{z}_j}(t, \hat{z}_1, ..., \hat{z}_{n-1}, \dot{\hat{z}}_1, ..., \dot{\hat{z}}_{n-1}) \dot{c}_j. \end{aligned} \tag{B2}$$

But the true rigid solution certainly minimizes Eq. (3.7) as it sets it to 0, thus being a solution of Eq. (3.8). It follows that the first term on the righthand side of Eq. (B2) is 0, from which Eq. (3.10) is concluded.

Appendix C

In this appendix we indicate the method by which equations of the type 3.14 and 3.15 were derived.

Equation 3.14 is derived by taking the necessary partial derivatives of D (Eq. (3.7)) by z_i and $\dot{z}_i$ and combining them as indicated in Eq. (3.11). These derivatives are evaluated at the true solution itself, that is at $\hat{z}_i(t)$ and $\dot{\hat{z}}_i(t)$, and because we only considered rigid motions in this paper, we could use the following relationship to simplify the results:

$$a_{ij} + (\hat{z}_i - \hat{z}_j)(\dot{\hat{z}}_i - \dot{\hat{z}}_j) = 0. \tag{C1}$$

Eq. (C1) is derived by setting the distances between points i and j as constant and taking the temporal derivative of this distance.

The conclusion of Eq. (3.15) is derived by noting that, in a rigid rotation around an axis perpendicular to the viewing axis, with constant angular velocity ω, $\hat{z}_i$ can be written as:

$$\hat{z}_i(t) = e_i cos(\omega t + \phi_i). \tag{C2}$$

Then, by direct substitution of Eq. (C2) in Eq. (3.14) and its integration as indicated in Eq. (3.15), we derive the stated result.

The calculations explained in this appendix, though straightforward, are cumbersome, and were done by using Macsyma, a computer system for performing algebraic manipulation.

Appendix D

In this appendix we show that if $\mathbf{C}$ is an $n \times n$ matrix over an algebraically closed field then:

$$\rho^j(\mathbf{C}) = \rho(\mathbf{C}^j), \tag{D1}$$

where ρ is the spectral radius.

Let $\mathbf{P}$ be such that:

$$\mathbf{P}^{-1}\mathbf{C}\mathbf{P} = \mathbf{J}, \tag{D2}$$

where $\mathbf{J}$ is in Jordan canonical form. It follows that:

$$\mathbf{C}^j = \mathbf{P}\mathbf{J}^j\mathbf{P}^{-1}. \tag{D3}$$

Similar matrices have similar characteristic values and thus similar spectral radii, thus:

$$\rho(\mathbf{C}) = \rho(\mathbf{J}), \tag{D4}$$

and

$$\rho(\mathbf{C}^j) = \rho(\mathbf{J}^j). \tag{D5}$$

Direct inspection of the matrices $\mathbf{J}^j$ show that they have the same type of triangularization as $\mathbf{J}$ (that is, upper or lower triangularization), with the elements in the diagonal being the $j-th$ power of those of $\mathbf{J}$. Thus it follows that:

$$\rho(\mathbf{J}^j) = \rho^j(\mathbf{J}), \tag{D6}$$

which together with Eqs. (C4) and (C5) imply the result stated in Eq. (D1).