The Sensitivity Of Motion and Structure Computations

John L. Barron, Allan D. Jepson* and John K. Tsotsos*

Department of Computer Science
University of Toronto
Toronto, Canada, M5S 1A4

Abstract

We address the problem of interpreting image velocity fields generated by a moving monocular observer viewing a stationary environment under perspective projection to obtain 3-D information about the relative motion of the observer (egomotion) and the relative depth of environmental surface points (environmental layout). The algorithm presented in this paper involves computing motion and structure from a spatio-temporal distribution of image velocities that are hypothesized to belong to the same 3-D planar surface. However, the main result of this paper is not just another motion and structure algorithm that exhibits some novel features but rather an extensive error analysis of the algorithm's preformance for various types of noise in the image velocities.

Waxman and Ullman [83] have devised an algorithm for computing motion and structure using image velocity and its 1^{st} and 2^{nd} order spatial derivatives at one image point. We generalize this result to include derivative information in time as well. Further, we show the approximate equivalence of reconstruction algorithms that use only image velocities and those that use one image velocity and its 1^{st} and/or 2^{nd} spatio-temporal derivatives at one image point. The main question addressed in this paper is: "How accurate do the input image velocities have to be?" or equivalently, "How accurate does the input image velocity and its 1^{st} and 2^{nd} order derivatives have to be?". The answer to this question involves worst case error analysis. We end the paper by drawing some conclusions about the feasibility of motion and structure calculations in general.

1.1 Introduction

In this paper, we present a algorithm for computing the motion and structure parameters that describe **egomotion** and **environmental layout** from image velocity fields generated by a moving monocular observer viewing a stationary environment. Egomotion is defined as the motion of the observer relative to his environment and can be described by 6 parameters; 3 depth-scaled translational parameters, $\vec{u}$, and 3 rotation parameters, $\vec{\omega}$. Environmental layout refers to the 3-D shape and location of objects in the environment. For monocular image sequences, environmental layout is described by the normalized surface gradient, $\vec{\alpha}$, at each image point. To determine these motion and structure parameters we derive nonlinear equations relating image velocity at some image point $\vec{Y}(\vec{P}',t')$ to the underlying motion and structure parameters at $\vec{Y}(\vec{P},t)$. The computation of egomotion and environmental layout from image velocity is sometimes called the **reconstruction** problem; we reconstruct the observer's motion, and the layout of his environment, from (time-varying) image velocity. A lot of research has been devoted to devising reconstruction algorithms. However, a little addressed issue concerns their performance for noisy input: how accurate does the input image velocity have to be to get useful output?

1.2 Previous Work

The most common approach to monocular reconstruction involves solving (generally nonlinear) systems of equations relating image velocity (or image displacement) to a set of motion and structure parameters ([Longuet-Higgins 81], [Tsai and Huang 84], [Prazdny 79], [Roach and Aggarwal 80], [Webb and Aggarwal 81], [Fang and Huang 84a,b], [Buxton et al 84], [Williams 80], [Dreschler and Nagel 82], [Lawton 83]). Some of the issues that arise for these algorithms are the need for good initial guesses of the solutions, the possibility of multiple solutions and the need for accurate input. The latter is by far the most important issue if the reconstruction approach is to be judged a success. As Waxman and Ullman [85] and others have noted, reconstruction techniques that use image velocities of neighbouring image points require accurate differences of these similar velocities. That is, solving systems of equations effectively requires subtraction of very similar quantities: the error in the quantities themselves may be quite small but since the magnitudes of these differences are quite small, the relative error in the differences can be quite large. Hence, such techniques can be expected to be sensitive to input errors.

A second approach to reconstruction involves solving nonlinear systems of equations relating local image velocity information (one image velocity and its 1^{st} and 2^{nd} order spatial derivatives) to the underlying motion and structure parameters ([Longuet-Higgins and Prazdny 80], [Waxman and Ullman 85]). The rationale is that using local information about one image point means that the problem of similar neighbouring image velocities can be averted. However, this is replaced with the problem of computing these 1^{st} and 2^{nd} order spatial derivatives. Waxman and Wohn [85] propose that these derivatives be found by solving linear systems of equations: where each equation specifies the normal component of image velocity on a moving non-occluding contour in terms of a Talyor series expansion of the x and y components of image velocity. In effect, their reconstruction algorithm divides the computation into two steps: use a normal velocity distribution to compute image velocity and its 1^{st} and 2^{nd} order spatial derivatives at an image point and then use these as input to an algorithm that solves the non-linear equations relating motion and structure to the image velocity and its 1^{st} and 2^{nd} order derivatives.

Only recently, have researchers begun to address the use of temporal information, such as temporal derivatives, in reconstruction ([Subbarao 86], [Bandyopadyay and Aloimonos 85]). We note that others' use of temporal derivative information and our use of time-varying image velocities are effectively equivalent; image velocity fields (at least locally) can be derived from one image velocity and its 1^{st} and/or 2^{nd} spatial and temporal derivatives and vice-versa. Indeed, image velocity fields are often used in the derivation of spatial and temporal image velocity information.

It is somewhat disappointing that almost none of these reconstruction techniques have been successfully applied to flow fields calculated from realistic scenes. Primarily, the problem is the difficulty in computing accurate flow fields. There has been little or no error analysis in previous monocular reconstruction work, although some researchers, such as [Waxman and Ullman 85], [Buxton et al 84], [Aloimonos and Rigoutsos 86], [Snyder 86] and [Subbarao 86] have begun to consider the inherent sensitivity of their algorithms to random noise in the input. See [Barron 84,87] for a more detailed survey of reconstruction techniques and their problems.

1.3 Underlying Assumptions

In order to relate a spatio-temporal distribution of image velocities to the motion and structure parameters at some image point we need to make some assumptions:

(a) 3-D objects are assumed to be rigid. The **rigidity** assumption ensures that the image velocity of an object's point is due entirely to the point motion with respect to the observer and not due to changes in the object's shape.

(b) The 3-D surfaces of objects can be described locally as a plane. The **local planarity** assumption means curved surfaces are treated as collections of adjacent planes.

(c) The observer rotates with a constant angular velocity for some small time interval. Webb and Aggarwal [81] call this the **fixed axis** assumption.

(d) The spatio-temporal distribution of image velocity results from

* Also, Canadian Institute for Advanced Research

3-D points on the same planar surface. We call this the **same surface** assumption.

(e) The observer translates with a constant speed (and possibly constant acceleration).

The use of a spatio-temporal distributions of image velocity means motion and structure is computed using local spatio-temporal data; thus we are not necessarily restricted to stationary environments as we can treat each independently moving surface as stationary relative to the moving observer. Similarly, the constraints on the motion need only be satisfied for short time intervals.

In [Barron et al 87a] we treated violation of these assumptions as one type of error in the input data. The use of the local planarity and fixed axis assumptions means that the point-to point correspondence problem does not have to be solved, i.e. we do not have to use velocities of the same 3-D points at different time intervals, as it is now mathematically possible to relate image velocities distributed in space and time at any point and time on a 3-D planar surface to the motion and structure parameters of any other point on the planar surface at any other time (where these assumptions are reasonably satisfied) [1]. Other researchers, such as [Kanatani 85] and [Aloimonos and Rigoutsos 86], have also advocated a correspondence-less approach. The computation of image velocity may require the solving of correspondence although there is a group of techniques based on the relationship between spatial and temporal grayvalue gradients, for example, [Horn and Schunck 81], for determining image velocity without the need to compute correspondence.

1.4 Previous Results

In a previous paper [Barron et al 87a] we presented a first set of results for the monocular algorithm presented in this paper. Some of the more important ones are:

(1) The use of a spatio-temporal distribution of image velocity rather than a purely spatial distribution of image velocity generally reduced the amplification of input to output error. As well, increasing the spatial extent of the image points where image velocity are measured also reduced error amplification.

(2) It appears that the accuracy with which image velocities can be computed is much more important that the satisfaction of the various assumptions. The solutions are not especially sensitive to small violations of the assumptions.

(3) The error in the initial guess (required for Newton's method) can be quite large (100% and more) and convergence can still be obtained for most cases when image velocity error is present.

(4) For exact image velocities, we found multiple solutions even though theoretical results suggested unique solutions. The analysis showed that it is possible for 2 distinct sets of flow fields to have 4 image velocities in common.

(5) We conducted a best, random and worst case error analysis for a related set of motion and structure parameters. (The 3 error types were scaled for comparison purposes.) The difference between the best, random and worst case results was significant. This suggests that tests based on a random noise alone are inadequate.

(6) The use of time allowed us to analyze motion and structure combinations that were singular at one time. For example, the motion and structure: $\vec{U}$=(0,0,1000), $\vec{\alpha}$=(0,0,1) and $\vec{\omega}$=(0.2,0,0) is unanalyzable at time 0 but can be solved given image velocities distributed over a short time interval.

We have also devised a binocular reconstruction [Barron et al 87b] that contains the monocular algorithm presented in this paper as a special case.

1.5 Contributions of this Paper

The algorithm presented in this paper involves solving nonlinear system of equations that relate a spatio-temporal distribution of image velocity to a set of motion and structure parameters at some image point at a particular time. We conduct a Taylor series expansion analysis and show the equivalence of using a mean image velocity and its 1^{st} and/or 2^{nd} order spatio-temporal derivatives to using 4 individual image velocities when a moving monocular observer views a planar surface. The main results of the paper are contained in the error analysis. In particular:

(1) We conduct a scaled best, random and worst case error analysis for a set of related motion and structure combinations when image velocities are used as input. (The errors are scaled for comparison purposes.) We investigate the amplification of input errors for the whole solution and for $\vec{u}$, $\vec{\alpha}$ and $\vec{\omega}$ along with the relationship between worst case image velocity error and the error in the Taylor series expansion coefficients. We also investigate the algorithm's performance when there is a maximum of X% worst case error in any of the image velocities.

(2) We conduct best and worst case error analysis by adding worst case error directly to the Taylor series expansion coefficients. We are interested in how the algorithm performs when there is a maximum of X% worst case error in any of the Taylor series coefficient pairs. (In general, the worst case error directions for the image velocities and the Taylor series coefficients are different.)

2 Mathematical Preliminaries

In this section we present a brief description of our algorithm. Complete details are in [Barron 87].

2.1 Notation

We use notation $\vec{P}(t;\tau)$ to indicate a 3-D point measured at time t with respect to a coordinate system $\vec{X}(\tau)$. Similarly, $X_3(\vec{P},t;\tau)$ is the depth of $\vec{P}(t;\tau)$. $\vec{Y}(\vec{P},t)$ is the image of $\vec{P}(t;)$.

2.2 Physical Setup

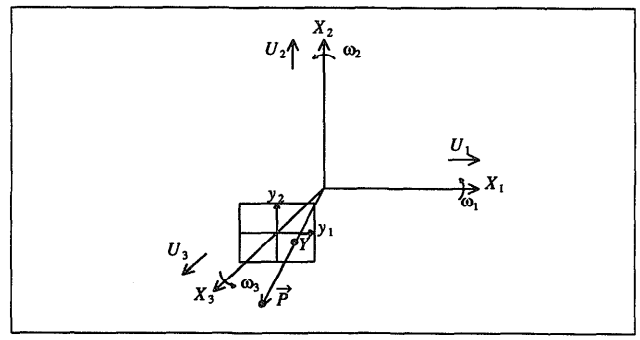

Figure 2.1 The Observer-Based Coordinate System. $\vec{U}=(U_1,U_2,U_3)$ is the observer's 3-D translational velocity while $\vec{\omega}=(\omega_1,\omega_2,\omega_3)$ is his 3-D rotational velocity. The image plane is at depth 1. The image of $\vec{P}$ is located at $\vec{Y}=(y_1,y_2,1)$. The origin of the image plane is (0,0,1). The X_3 axis is the line of sight.

We adopt a right-handed coordinate system as in Longuet-Higgins and Prazdny [80] which is shown in Figure 2.1. $\vec{U}=(U_1,U_2,U_3)$ is the translational velocity of the observer, centered at the origin of the coordinate system $\vec{X}(t)$ and $\vec{\omega}=(\omega_1,\omega_2,\omega_3)$ is the angular velocity of the observer.

2.3 The General Monocular Image Velocity Equation

We can write an equation relating image velocity at some image point $\vec{Y}(\vec{P}_j,t')$ to the monocular motion and structure parameters at some image point $\vec{Y}(\vec{P}_i,t)$ as

$$\vec{v}(\vec{Y}(\vec{P}_j,t'),t') = A_1(\vec{Y}(\vec{P}_j,t'))\, h(\vec{Y}(\vec{P}_i,t))\, Q_i(\vec{\omega},t,t')\, \vec{u}(\vec{Y}(\vec{P}_i,t),t;)$$
$$S_M(\vec{Y}(\vec{P}_i,t),\vec{Y}(\vec{P}_j,t'),t')\, T_M(\vec{Y}(\vec{P}_i,t),t,t')+A_2(\vec{Y}(\vec{P}_j,t'))\,\vec{\omega}(t;). \quad (2.3\text{-}1)$$

where $\vec{P}_i$ and $\vec{P}_j$ are 3-D points on the same planar surface and generally $\vec{Y}(\vec{P}_i,t)\neq\vec{Y}(\vec{P}_j,t')$. In the above equation

$$A_1(\vec{Y}(\vec{P},t)) = \begin{bmatrix} -1 & 0 & y_1 \\ 0 & -1 & y_2 \\ 0 & 0 & 0 \end{bmatrix} \quad (2.3\text{-}2)$$

and

(1) Of course, we must still be able to solve surface correspondence, i.e. we must be able to group together all image velocities distributed locally in space and time that belong to the same planar surface. See [Adiv 84] for one approach to this problem.

$$A_2(\vec{Y}(\vec{P},t)) = \begin{bmatrix} y_1y_2 & -(1+y_1^2) & y_2 \\ (1+y_2^2) & -y_1y_2 & -y_1 \\ 0 & 0 & 0 \end{bmatrix}, \quad (2.3\text{-}3)$$

$h(\vec{Y}(\vec{P},t))$ is the **perspective correction function** that specifies the ratio between the depth of $\vec{P}(t;t)$, $X_3(\vec{P},t;t)$ and its 3-D distance from the observation point, $||\vec{P}(t,t)||_2=(\vec{P}(t;t)\cdot\vec{P}(t;t))^{\frac{1}{2}}$, i.e.

$$h(\vec{Y}(\vec{P},t)) = \frac{||\vec{P}(t;t)||_2}{X_3(\vec{P},t;t)} = ||\vec{Y}(\vec{P},t)||_2 \quad (2.3\text{-}4)$$

and $\vec{u}(\vec{Y}(\vec{P},t),t;t)$ is the distance-scaled translational velocity of the observer,

$$\vec{u}(\vec{Y}(\vec{P},t),t;t) = \frac{\vec{U}(t;t)}{||\vec{P}(t;t)||_2}. \quad (2.4\text{-}5)$$

One of the advantages of using a single instantaneous image velocity field is that no assumptions about the observer's motion, for example his acceleration, have to be made. However, the use of a spatio-temporal distribution of image velocities requires that we relate the motion and structure parameters at one time to those at another time. Hence, we need to make assumptions about the observer's motion. In this paper, we consider two specific types of motion, although we emphasize that our treatment can be generalized to other motions as well. The two types of motion considered are:

Type 1: (Linear Motion, Rotating Observer): A vehicle is moving with constant translational velocity and has a camera mounted on it that is rotating with constant angular velocity.

Type 2: (Circular Motion, Fixed Observer): A vehicle with a fixed mounted camera is moving with constant translational and angular velocity.

$Q_1(\vec{\omega},t,t')=R^T(\omega,t')R(\vec{\omega},t)$ and $Q_2(\vec{\omega},t,t')=I$ (the identity matrix), for Types 1 and 2 motion respectively. $R(\vec{\omega},t)$ is an orthogonal matrix specifying the rotation $||\vec{\omega}||_2 t$ of $\vec{X}(t)$ with respect to $\vec{X}(0)$. S_M, the **monocular spatial scaling function,**

$$S_M(\vec{Y}(\vec{P}_i,t),\vec{Y}(\vec{P}_j,t),t) = \frac{\vec{\alpha}(\vec{P}_j,t;t)\cdot\vec{Y}(\vec{P}_j,t)}{\vec{\alpha}(\vec{P}_i,t;t)\cdot\vec{Y}(\vec{P}_i,t)} = \frac{X_3(\vec{P}_i,t;t)}{X_3(\vec{P}_j,t;t)}. \quad (2.3\text{-}6)$$

specifies the depth ratio of two 3-D points, $\vec{P}_i$ and $\vec{P}_j$ on the same planar surface at the same time. The **monocular temporal scaling function,**

$$T_M(\vec{Y}(\vec{P}_i,t),t,t') = \frac{X_3(\vec{P}_i,t;t)}{X_3(\vec{P}_k,t';t')} \quad (2.3\text{-}7)$$

$$= \frac{\vec{\alpha}(\vec{P}_i,t'';t'')\cdot\vec{Y}(\vec{P}_i,t)}{\vec{\alpha}(\vec{P}_k,t'';t'')\cdot R^T(\vec{\omega},t'')R(\vec{\omega},t)\left[\vec{Y}(\vec{P}_i,t)-\Delta\vec{d}(\vec{P}_i,t,t'';t)h(\vec{Y}(\vec{P}_i,t))\right]}$$

specifies the depth ratio of two 3-D points, $\vec{Y}(\vec{P}_i,t)=\vec{Y}(\vec{P}_k,t'')$

In special cases, equation (2.3-1) reduces to either a purely spatial or a purely temporal image velocity equation when $S_M=1$ or $T_M=1$. Given eight distinct components of image velocity distributed over space and time, but on the same 3-D planar surface, we can construct and solve a non-linear system of equations to determine the motion and structure parameters.

2.4 The Non-Uniqueness of the Solutions

Because we are solving non-linear systems of equations we need to be concerned about the uniqueness of our solution. Hay [66] was the first to investigate the inherent non-uniqueness of the visual interpretation of a stationary planar surface. He showed that for any planar surface there are at most two sets of motion and structure parameters that give rise to the same image velocity field for that surface. Hay also showed that given two views of such a surface only one unique set of motion and structure parameters was capable of correctly describing the image velocity field. Waxman and Ullman [83] carried this result one step further by showing the **dual** nature of the solutions: given one set of motion and structure parameters it is possible to derive a second set in terms of the first analytically. If this second solution is then substituted back into the equations specifying the duality, the first solution is obtained. Given one set of motion and structure parameters, $\vec{u}_1$, $\vec{\alpha}_1$ and $\vec{\omega}_1$, we can derive expressions for the dual solution, $\vec{u}_2$, $\vec{\alpha}_2$ and $\vec{\omega}_2$, at some point $\vec{Y}(\vec{P},t)$ as

$$\vec{u}_2(\vec{Y}(\vec{P},t),t;t) = \vec{\alpha}_1(\vec{P},t;t)\frac{\vec{u}_1(\vec{Y}(\vec{P},t),t;t)\cdot\vec{Y}(\vec{P},t)}{\vec{\alpha}_1(\vec{P},t;t)\cdot\vec{Y}(\vec{P},t)}, \quad (2.4\text{-}1a)$$

$$\vec{\alpha}_2(\vec{P},t;t) = \frac{\vec{u}_1(\vec{Y}(\vec{P},t),t;t)}{||\vec{u}_1(\vec{Y}(\vec{P},t),t;t)||_2} \quad (2.4\text{-}1b)$$

and

$$\vec{\omega}_2(t;t) = \vec{\omega}_1(t;t) + \begin{bmatrix} 0 & -\alpha_{13}(\vec{P},t;t) & \alpha_{12}(\vec{P},t;t) \\ \alpha_{13}(\vec{P},t;t) & 0 & -\alpha_{11}(\vec{P},t;t) \\ -\alpha_{12}(\vec{P},t;t) & \alpha_{11}(\vec{P},t;t) & 0 \end{bmatrix} \frac{\vec{u}_1(\vec{Y}(\vec{P},t),t;t)h(\vec{Y}(\vec{P},t))}{\vec{\alpha}_1(\vec{P},t;t)\cdot\vec{Y}(\vec{P},t)}. \quad (2.4\text{-}1c)$$

In (2.4-1c) we use the notation $\vec{\alpha}_1=(\alpha_{11},\alpha_{12},\alpha_{13})$ in (2.4-1c). Obviously, when $\vec{\alpha}(\vec{P},t;t)=\frac{\vec{u}(\vec{Y}(\vec{P},t),t;t)}{||\vec{u}(\vec{Y}(\vec{P},t),t;t)||_2}$, the solution is unique as the dual solution reduces to the first solution. Subbarao and Waxman [85] have also showed the uniqueness of the motion and structure parameters over time as well.

These theoretical results suggest that the possibility of multiple (non-dual in the spatial case) solutions is non-existent. However, they only hold when the whole flow field is considered. It is possible for two distinct image velocity fields to have four common image points at four times with the same the image velocity values. Hence, the analysis of the four image velocities may give rise to any of the sets of motion and structure parameters having those four image velocities in common. An example of such a situation is shown in [Barron et al 87a].

2.5 Singularity

If $\vec{u}=(0,0,0)$ then the system of equations is singular. In fact, when $\vec{u}\ll\vec{\omega}$, its condition number becomes very large; very small input error causes instability in the solution technique. Also, Fang and Huang [84a] and others have shown that the solution does not exist using the image velocities at three or more collinear 3-D points (as the determinant of the J is 0). We have also observed that the solution using two image velocities at each of two 3-D points on the same planar surface at two distinct times cannot be obtained. The motion of 2 points can be caused by an infinite number of motion and structure combinations. As well, there are particular motion and structure combinations that cannot be recovered at one time. For example, if $\vec{U}=(0,0,a)$, $\vec{\alpha}=(0,0,1)$ and $\vec{\omega}=(0,b,0)$ at time 0, then the values of constants a and b can be arbitrarily set to yield the same image velocity field; hence, it is impossible to distinguish the translational and rotational components of image velocity from each other. Other conditions of singularity are under active investigation.

3 Experimental Technique

In this section we discuss the implementation of our algorithm and present details of our sensitivity analysis.

3.1 Implementation

Newton's method is used to solve the systems of non-linear equations. Since only 2 components of $\vec{\alpha}$ are independent, i.e. $||\vec{\alpha}||_2=1$, we add extra row to the Jacobian matrix, J to ensure the computed $\vec{\alpha}$ is normalized; hence J is a full rank 9 matrix. The 9^{th} value of $\vec{f}_M$, the measured image velocities is then set to 1.

When $\vec{\omega}$ is known to be zero, i.e. in the case of pure translation (Type 1 and Type 2 motions are equivalent here) we solve a 6×6 Jacobian instead of a 9×9 one. We compute a 9×6 Jacobian (the 3 columns corresponding to $\vec{\omega}$ are not computed). We let the LU decomposition of J choose the bests 6 rows of J, with the provision that the normalization row is always one of the chosen rows.

3.2 Sensitivity Analysis

We compute an error vector, $\Delta\vec{f}_M$, which when added to $\vec{f}_M$, yields the perturbed input, $\vec{f}_M'=\vec{f}_M+\Delta\vec{f}_M$. For X% **random** case error, we compute four random 2-component unit vectors, $\hat{n}_j$, $j=1,...,4$, and then compute each i^{th} component of $\Delta\vec{f}_M$ as

$$\begin{bmatrix} \Delta f_{iM} \\ \Delta f_{i+1\,M} \end{bmatrix} = \frac{X}{100}\hat{n}_j||\vec{v}_j||_2, j=1,...,4, \; i=j\times 2-1. \quad (3.2\text{-}1)$$

Thus X% random error is added to each image velocity. $\Delta f_{9,M}$ is 0, i.e. we do not add error to the normalization constant. Using $\Delta\vec{f}_M$ for random error we compute $\Delta f_{norm} = ||\Delta\vec{f}_M||_2$. We use forward and inverse [2] iteration on J to compute normalized **best** and **worst** case error directions, $\hat{e}_B$ and $\hat{e}_W$. We compute $\Delta\vec{f}_M = \hat{e}_B \Delta f_{norm}$ as X% scaled best case image velocity error and $\Delta\vec{f}_M = \hat{e}_W \Delta f_{norm}$ as X% **scaled worst case** image velocity error. In either case $\Delta\vec{f}_M$ is scaled to the same size as the random $\Delta\vec{f}_M$ for comparison purposes. When $\vec{\omega}$ is known to be 0 the appropriate 6×6 Jacobian is used in the forward and inverse iteration calculation. We can also compute X% **relative worst case** image velocity error by computing $\Delta\vec{f}_M$ so that the image velocity with the largest error is X%.

We perform a Taylor series expansion of image velocity. In the spatial case, we can write

$$\vec{f}_M \approx A_s \vec{g}_s \tag{3.2-2}$$

for small spatial extents. Here $\vec{g}_s$ is given as

$$\vec{g}_s = \left[v_{m1}, v_{m2}, \frac{\partial v_1}{\partial y_1}, \frac{\partial v_2}{\partial y_1}, \frac{\partial v_1}{\partial y_2}, \frac{\partial v_2}{\partial y_2}, \frac{\partial^2 v_1}{\partial y_1 \partial y_2}, \frac{\partial^2 v_2}{\partial y_1 \partial y_2} \right]$$
$$= (\vec{g}_1, \vec{g}_2, \vec{g}_3, \vec{g}_4) \tag{3.2-3}$$

and A_s is given as

$$A_s = \begin{bmatrix} I_2 & I_2\Delta y_{11} & I_2\Delta y_{12} & \begin{bmatrix} \Delta y_{11}^2 & \Delta y_{11}\Delta y_{12} \\ \Delta y_{11}\Delta y_{12} & \Delta y_{12}^2 \end{bmatrix} \\ I_2 & I_2\Delta y_{21} & I_2\Delta y_{22} & \begin{bmatrix} \Delta y_{21}^2 & \Delta y_{21}\Delta y_{22} \\ \Delta y_{21}\Delta y_{22} & \Delta y_{22}^2 \end{bmatrix} \\ I_2 & I_2\Delta y_{31} & I_2\Delta y_{32} & \begin{bmatrix} \Delta y_{31}^2 & \Delta y_{31}\Delta y_{32} \\ \Delta y_{31}\Delta y_{32} & \Delta y_{32}^2 \end{bmatrix} \\ I_2 & I_2\Delta y_{41} & I_2\Delta y_{42} & \begin{bmatrix} \Delta y_{41}^2 & \Delta y_{41}\Delta y_{42} \\ \Delta y_{41}\Delta y_{42} & \Delta y_{42}^2 \end{bmatrix} \end{bmatrix} . \tag{3.2-4}$$

For the spatio-temporal case, we can write

$$\vec{f}_M \approx A_t \vec{g}_t \tag{3.2-5}$$

for small spatio-temporal extents. Here

$$\vec{g}_t = \left[v_{m1}, v_{m2}, \frac{\partial v_1}{\partial y_1}, \frac{\partial v_2}{\partial y_1}, \frac{\partial v_1}{\partial y_2}, \frac{\partial v_2}{\partial y_2}, \frac{\partial v_1}{\partial t}, \frac{\partial v_2}{\partial t} \right]$$
$$= (\vec{g}_1, \vec{g}_2, \vec{g}_3, \vec{g}_4) \tag{3.2-6}$$

and

$$A_t = \begin{bmatrix} I_2 & I_2\Delta y_{11} & I_2\Delta y_{12} & I_2\Delta t_1 \\ I_2 & I_2\Delta y_{21} & I_2\Delta y_{22} & I_2\Delta t_2 \\ I_2 & I_2\Delta y_{31} & I_2\Delta y_{32} & I_2\Delta t_3 \\ I_2 & I_2\Delta y_{41} & I_2\Delta y_{42} & I_2\Delta t_4 \end{bmatrix} . \tag{3.2-7}$$

We can compute the Taylor series expansion using both perfect $\vec{f}_M$ and noisy $\vec{f}_M'$, to get $\vec{g}'$ and $\vec{g}$ respectively. $\Delta\vec{g}$ is simply $\vec{g}'-\vec{g}$.

We compute X% **relative best case** and X% **relative worst case** error in $\Delta\vec{g}$ by performing forward and inverse iteration on $A^{*-1}J$ where A^* is an 9×9 matrix computed using $A=A_s$ or $A=A_t$ as $A^* = \begin{bmatrix} A & 0_C \\ 0_R & 1 \end{bmatrix}$, where 0_R and 0_C are 8 component row and column vectors containing all 0's and $\vec{g}^*$ is simply $(\vec{g},1)$. These best and worst directions are then scaled so that there is a maximum of X% in any of the Taylor series coefficients, $\Delta\vec{g}_i$, i=1,...,4 where $\Delta\vec{g}_i$ is the error in $\vec{g}_i$. When $\vec{\omega}$ is zero, we cannot conduct this error analysis as the 9×9 Jacobian is singular.

Output error in the computed solution $\vec{s}$ and in its components, $\vec{u}$, $\vec{\alpha}$ and $\vec{\omega}$ are computed simply as the L_2 norm of their difference from the correct solution/component over the L_2 norm of the correct solution/component. In purely spatial cases, we also compute the dual of $\vec{s}$, $\vec{s}_D$, using (3.4) and compute the output error as the minimum of error in $\vec{s}$ or $\vec{s}_D$. Since $\vec{\alpha}'$ and $-\vec{\alpha}'$ specify the same surface gradient, we always "flip" $\vec{\alpha}'$ before the output error is calculation if the flipped $\vec{\alpha}'$ is closer to $\vec{\alpha}$ than the original $\vec{\alpha}'$ i.e. $||\vec{\alpha}+\vec{\alpha}'||_2 < ||\vec{\alpha}-\vec{\alpha}'||_2$. The error in the various $\Delta\vec{g}_i$, i=1 to 4 and $\Delta\vec{g}$ is simply computed as the L_2 norm of the difference over the L_2 norm of the correct value.

4 Experimental Results

We use the motion and structure described in the table below for the experimental results presented in this paper.

Motion	$\vec{U}$	$\vec{\alpha}$	$\vec{\omega}$	Type
1	0,0,1000	0,0,1	0,0,0	1,2
2	0,0,1000	0,0,1	0,0.2,0	1
3	0,0,1000	0,0,1	0,0.2,0	2

X_3 is 2000 in all cases. Image coordinates are measured on a 256×256 display device and so pixel coordinates are scaled by 256 to produce the corresponding f coordinates. Thus the solution point, $\vec{Y}_s$, is (20,20) in pixels or (0.078125,0.078125,1) in f units. For a temporal extent 0–t we measure image velocities at the following image point offsets and times

Δy_1	Δy_2	Time
50	50	0.0
-50	50	t/3
-50	-50	2t/3
50	-50	t

where $\vec{Y}_i = \vec{Y}_s + (\Delta y_{i1}, \Delta y_{i2}, 0)$, i=1 to 4. The viewing angle of these points, which we call their **spatial extent**, is computed as the maximum diagonal angle subtended by the points, i.e. 33.05°. Temporal extent 0–t is varied by varying t from 0 to 1 or 0.3 to 1 in 8 equal steps; a temporal extent of 0-0 corresponds to the purely spatial case. (As we have already pointed out, when $\vec{\omega}$=(0,0.2,0), the motion is singular for temporal extent 0-0 but is can be solved at other temporal extents. When $\vec{\omega}$ is known to be zero we can solve this motion for temporal extent 0-0 provided we enforce $\vec{\omega}=\vec{0}$.) Image velocity error is varied from 0-1.4% in 0.2% steps while Taylor series cofficient error is varied from 0-14% in 2% steps.

In the first experiment we vary image velocity error against temporal extent. Tables 1, 2 and 3 shows the overall amplification factors [3], their standard deviations and the number of solved runs when the image velocity error was not 0% (out of a maximum of 56) computed for the 3 motions for $\vec{s}$, $\vec{u}$, $\vec{\alpha}$ and $\vec{\omega}$ for all solved runs. Best case results are quite good, especially when compared with the corresponding random and worst case results. Random output is about ½ the worst case output error. We include random results only to show the inadequacy of an error analyisis that only involves performing a few runs with random noise in the input. Unless a particular run is made n times (n a sufficiently large integer) for random noise in the input we cannot draw any useful conclusions. The largest output error for these n runs should approach worst case results while the average output error for the n runs comprises average case results. Table 4 shows the overall amplification factors for the 3 motions when 0-1.4% relative worst case image velocity error is used instead. These results indicate that worst case error of 1.4% and smaller can produce unusable output. It seems that we need image velocity measurements to be quite accurate.

An examination of the error velocity means and differences for the above runs confirms the hypothesis about image velocity error presented in section 1.2. For best case results, the error in the means is larger than the error in the differences while for worst case results the error in the differences is larger. In another experiment (see [Barron 87]) we added worst case error directly to the means and differences of the image velocities. The results further confirm the hypothesis as worst case mean error produced very small error amplification while worst case difference error produced much larger ones. Indeed, even when large worst case mean error was used (up

(2) We note that the best and worst directions so calculated are for the initial linear system of equations, $J\vec{h}_0=\vec{f}_0$. It is possible that the best and worst directions for the nonlinear system of equations are different, although we expect these directions to be quite close to the computed best and worst directions for small input errors.

(3) Overall amplification factors are computed as the average of output error divided by input error for all solved runs where input error is not 0%.

to 49% error in image velocities) worst case mean amplification factors were still less than 1. In fact, worst case mean error results are almost as good as best case error results.

The second experiment involves adding relative best and worst case error to the Taylor coefficients, $\vec{g}_i$, and then computing motion and structure from the resulting image velocity fields. Preliminary results are presented in Table 5 for the 2nd and 3rd motions. The standard deviations are much larger here. This is because the magnitude of the input error increases significantly in time; hence, the output error also increases in time. For the smaller temporal extents the amplification is typically 2-3. Again we note that mean error (in $\vec{g}_1$) is larger in best case results than in worst case results while derivative error (in $\vec{g}_2$, $\vec{g}_3$ and $\vec{g}_4$ is larger in worst case results than in best case results. These and earlier results suggest that adding error to velocity means has only a minimal effect on the error amplification but that adding error to the velocity differences/derivatives has a much greater effect on the error amplification.

It is interesting to note the relationship between image velocity error and error in the Taylor coefficients for that same image velocity field. Table 6 shows the error in $\vec{g}_1$, $\vec{g}_2$, $\vec{g}_3$, $\vec{g}_4$ and $\vec{g}$ for temporal extents 0–0.3 and 0–1 for the 3 motions when 1.4% scaled worst case image velocity error is present in the input. Its seems that error in $\vec{g}_4$ is by far the largest. Overall the error in $\vec{g}$ is 2-3 times the error in the image velocities (4-6 times if we look at the L_2 norm error in the input) while the error in the various $\vec{g}_i$ can be 10-15 times larger. [Waxman and Ullman 85] note that in the spatial case recovery of motion and structure when there is large error in $\vec{g}_4$ is quite robust.

Changing $\vec{V}_i$ to (0,0) from (20,20), we conduct a last set of tests for the 3rd motion, varying spatial extent (the magnitudes of the coordinates of the four corners on the square) to have values 1^o, 14^o and 70^o (the full image). We vary t from 0.3 to 1 for these tests. Tables 7 and 8 show the overall amplification factors for these tests. We can solve most of the runs, even when the spatial extent is only 1^o and the relative error is 14%. Of course, the corresponding image velocity error is quite small. As expected, best case results are quite good. On the other hand worst case results are not nearly as good. Large amplification factors means that (effectively) the output is not useful even when a solution is obtained as is the case for most of the runs. Again, the magnitude of the actual image velocity error increased with time and is entirely due to the error in $\vec{g}_4=\frac{\partial \vec{v}}{\partial t}$. The errors in $\vec{g}_1$, $\vec{g}_2$ and $\vec{g}_3$ are quite small relative to this error. When the spatial extent is either 14^o or 70^o we observe an improvement in time.

5 Conclusions

The results of the above experiments suggest that reconstruction algorithms that use individual image velocities need them to within 1% or better or equivalently those algorithms that use local image velocity information (for example, [Waxman and Ullman 83]) need their input to within 10% accuracy. Derivative information is usually calculated directly from velocity fields, for example by using a least squares computation to related normal image velocity to the $\vec{g}_i$'s (see Waxman and Ullman [83]). Current techniques for measuring image velocity cannot produce this required accuracy. This may appear to call into question the feasibility of the reconstruction approach. However, an alternative approach is suggested from the experimental results:

(1) Compute one mean image velocity that corresponds to $\vec{g}_1 \Rightarrow \vec{v}_m$ (as we have seem the error in this velocity can be quite large) using one of the many conventional image velocity measurement techniques available (for example [Horn and Schunk 81] or [Barnard and Thompson 80]).

(2) Use separate techniques to measure spatio-temporal derivative information directly from raw time-varying intensity data. It may well be that such techniques will be able to measure the derivative information within the required accuracy. We advocate the design of such measurement techniques as a new research area.

The derivative data could then be used directly in the motion and structure calculation (such as Waxman and Ullman's algorithm) or first be converted into time-varying image velocity fields which, in turn, are used as input to, say, the algorithm presented in this paper.

Bibliography

(1) Adiv G., 1984, "Determining 3-D Motion and Structure from Optical Flow Generated by Several Moving Objects", COINS Technical Report 84-07, University of Massachusetts, April.

(2) Aloimonos J.Y. and I. Rigoutsos, 1986, "Determining the 3-D Motion of a Rigid Planar Patch Without Correspondence, Under Perspective Projection, Proc. Workshop on Motion: Representation and Analysis, May 7-9.

(3) Bandyopadhyay A. and J.Y. Aloimonos, 1985, "Preception of Rigid Motion from Spatio-Temporal Derivatives of Optical flow", TR 157, Dept. of Computer Science, University of Rochester, NY, March.

(4) Barnard S.T. and W.B. Thompson, 1980, "Disparity Analysis of Images", PAMI-2, No. 4, July, pp333-340.

(5) Barron, J.L., 1984, "A Survey of Approaches for Determining Optic Flow, Environmental Layout and Egomotion", RBCV-TR-84-5, Dept. of Computer Science, University of Toronto, November.

(6) Barron, J.L., 1987, "Determination of Egomotion and Environmental Layout From Noisy Time-Varying Image Velocity in Monocular and Binocular Image sequences", forthcoming PhD thesis, Dept. of Computer Science, University of Toronto.

(7) Barron J.L., A.D. Jepson and J.K. Tsotsos, 1987a, "Determining Egomotion and Environmental Layout From Noisy Time-Varying Image velocity in Monocular Image sequences", submitted for publication.

(8) Barron J.L., A.D. Jepson and J.K. Tsotsos, 1987b, "Determining Egomotion and Environmental Layout From Noisy Time-Varying Image velocity in Binocular Image sequences", IJCAI87, August, Milan, Italy.

(9) Buxton B.F, H. Buxton, D.W. Murray and N.S. Williams, 1984, "3-D Solutions to the Aperture Problem", in **Advances in Artificial Intelligence**, T. O'Shea (editor), Elsevier Science Publishers B.V. (North Holland), pp631-640.

(10) Dreschler L.S. and H.-H. Nagal, 1982, "Volumetric Model and 3-D Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a Street Scene", CGIP 20, pp199-228.

(11) Fang J.-Q. and T.S. Huang, 1984a, "Solving Three-Dimensional Small-Rotation Motion Equations: Uniqueness, Algorithms and Numerical Results", CVGIP 26, pp183-206.

(12) Fang J.-Q. and T.S. Huang, 1984b, "Some Experiments on Estimating the 3-D Motion Parameters of a Rigid Body from Two Consecutive Image Frames", PAMI, Vol.6, No.5, 1984, pp545-554.

(13) Hay J.C., 1966, "Optical Motions and Space Perception: An Extension of Gibson's Analysis", Psychological Review, Vol. 73, No. 6, pp550-565.

(14) Horn B.K.P. and B.G. Schunck, 1981, "Determining Optical Flow", AI 17, pp185-203.

(15) Kanatani K, 1985, "Structure from Motion without Correspondence: General Principle", Proc. IJCAI85, 886-888.

(16) Lawton D.T., 1983, "Processing Translational Motion Sequences", CGIP22, pp116-144.

(17) Longuet-Higgins H.C., 1981, "A Computer Algorithm for Reconstructing a Scene from Two Projections:, Nature 293, Sept., pp133-135.

(18) Longuet-Higgens H.C. and K. Prazdny, 1980, "The Interpretation of a Moving Image", Proc. Royal Society of London, B208, 1980, pp385-397.

(19) Prazdny K., 1979, "Motion and Structure From Optical Flow", IJCAI79, pp702-704.

(20) Roach J.W. and J.K. Aggarwal, 1980, "Determining the Movement of Objects from a Sequence of Images", PAMI, Vol. 2, No. 6, Nov., pp554-562.

(21) Snyder M.A., 1986, "The Accuracy of 3-D Parameters in Correspondence-Based Techniques: Startup and Updating", Proc. Workshop on Motion: Representation and Analysis, May 7-9.

(22) Subbarao M. and A.M. Waxman, 1985, "On the Uniqueness of Image Flow Solutions for Planar Surfaces in Motion", CAR-TR-114 (CS-TR-1485), Center for Automation Research, University of Maryland. (Also, 3rd Workshop on Computer Vision: Representation and Control, 1985.)

(23) Subbarao M., 1986, "Interpretation of Image Motion Fields: A Spatio-Temporal Approach", Proc. Workshop on Motion: Representation and Analysis, May 7-9.

(24) Tsai R.Y. and T.S. Huang, 1984, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces", IEEE PAMI, Vol. 6, No. 1, pp13-27.

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

(26) Waxman A.M. and S. Ullman, 1985, "Surface Structure and 3-D Motion from Image Flow Kinematics", Intl. Journal of Robotics Research, Vol.4, No.3, pp72-94.

(27) Waxman A.M. and K. Wohn, 1984, "Contour Evolution, Neighbourhood Deformation and Global Image Flow: Planar Surfaces in Motion", Intl. Journal of Robotics Research, Vol.4, No.3, pp95-108.

(28) Webb J.A. and J.K. Aggarwal, 1981, "Visually Interpreting the Motion of Objects in Space", IEEE Computer, Aug., pp40-46.

(29) Williams T.D., 1980, "Depth from Camera Motion in a Real World Scene", PAMI-2, No. 6, Nov., pp511-515.

Table 1: Scaled Best Case Error Amplification Factors

Motion	$\vec{s}$	St. Dev.	$\vec{u}$	St. Dev.	$\vec{\alpha}$	St. Dev.	$\vec{\omega}$	St. Dev.	Runs
1	0.116	0.003	0.210	0.006	0.065	0.003	-	-	56
2	0.148	0.003	0.236	0.005	0.011	0.0003	0.602	0.014	56
3	0.152	0.003	0.255	0.006	0.029	0.001	0.557	0.013	56

Table 2: Scaled Random Error Amplification Factors

Motion	$\vec{s}$	St. Dev.	$\vec{u}$	St. Dev.	$\vec{\alpha}$	St. Dev.	$\vec{\omega}$	St. Dev.	Runs
1	4.066	0.163	0.476	0.017	4.532	0.182	-	-	56
2	15.792	0.067	15.653	0.953	13.159	0.498	38.380	2.234	55
3	8.499	0.035	3.344	0.195	9.183	0.373	9.947	0.537	56

Table 3: Scaled Worst Case Error Amplification Factors

Motion	$\vec{s}$	St. Dev.	$\vec{u}$	St. Dev.	$\vec{\alpha}$	St. Dev.	$\vec{\omega}$	St. Dev.	Runs
1	7.672	0.261	0.545	0.017	8.563	0.291	-	-	56
2	55.982	2.547	82.004	4.539	22.084	0.901	191.256	10.331	45
3	19.798	0.491	5.699	0.195	21.957	0.540	18.316	0.559	56

Table 4: Relative Worst Case Error Amplification Factors

Motion	$\vec{s}$	St. Dev.	$\vec{u}$	St. Dev.	$\vec{\alpha}$	St. Dev.	$\vec{\omega}$	St. Dev.	Runs
1	4.064	0.117	0.291	0.008	4.536	0.133	-	-	56
2	18.706	0.719	20.432	1.057	14.788	0.530	49.910	2.443	48
3	13.792	0.357	4.195	0.152	15.260	0.391	13.357	0.431	56

Table 5 Amplification Factors for Taylor Series Coefficient Error

Motion	Best Case Error	St. Dev.	Runs	Worst Case Error	St. Dev.	Runs
2	0.093	0.046	56	7.955	3.244	33
3	0.179	0.086	56	5.072	2.136	54

Table 6 Error in Taylor Coefficients for 1.4% Scaled Worst Case Error

Motion	Temporal Extent	$\vec{g}_1$	$\vec{g}_2$	$\vec{g}_3$	$\vec{g}_4$	$\vec{g}$
1	0-0	0.10	2.25	1.13	0.32	0.98
1	0-1	1.162	0.81	1.21	1.32	1.19
2	0-0.3	2.27	2.60	4.07	17.55	4.86
2	0-1	1.93	2.57	4.41	28.61	3.80
3	0-0.3	3.66	2.10	5.69	21.12	6.99
3	0-1	3.29	1.47	5.28	7.20	4.12

Table 7: Amplification Factors for Relative Best Case Error

Spatial Extent	Amplification	St. Dev.	Runs
1^o	0.567	0.295	56
14^o	0.144	0.069	56
70^o	0.147	0.070	56

Table 8: Amplification Factors for Relative Worst Case Error

Spatial Extent	Amplification	St. Dev.	Runs
1^o	11.719	4.205	51
14^o	2.260	1.038	56
70^o	5.019	2.178	54