The Feasibility 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

The main contribution of this research is an extensive analysis of the noise sensitivity inherent in the motion and structure problem. We report results of an average and worst case error analysis for four types of image velocity input: full and normal image velocities and full and normal sets of image velocity and its derivatives. (These derivatives are simply the coefficients of a Taylor series expansion of image velocity about some point in space and time.) The main issues we address in this paper are: just how sensitive is motion and structure in the presence of noisy input, or alternately, how accurate do we need image velocity information, how much and what type of input data is needed and under what circumstances is motion and structure feasible? That is, when can we be sure that a motion and structure computation will produce usable results? We base our answers on a numerical error analysis we conduct for a large number of motions.

1.1 The Problem

In this paper we outline a motion and structure algorithm that uses time-varying image velocity information to compute the motion of a monocular observer moving relative to a stationary 3-D planar surface under perspective projection. Due to the depth-speed ambiguity experienced by a monocular observer, his translation $\vec{U}$ and the 3-D coordinates of environmental points on the planar surface, $\vec{P}$, cannot be recovered. We can only compute $\vec{u}$, the depth scaled observer translation and $\vec{\alpha}$, the normalized surface gradient of the 3-D planar surface. Since the rotational component of image velocity is independent of 3-D depth we can fully recover the observer's rotation, $\vec{\omega}$. To determine these motion and structure parameters we derive nonlinear equations relating image velocity and its derivatives at some image point and time, $\vec{Y}(\vec{P}',t')$ to the underlying motion and structure parameters at another image point and time, $\vec{Y}(\vec{P},t)$. Hence, we are able to use time-varying image velocity information in the computation of motion and structure. This is in sharp contrast with most previous approaches to motion and structure (see [Barron 84,88] and below for references) where image velocity information is measured at one time only. As we shall see below the use of time-varying image velocity information can often reduce error sensitivity in motion and structure calculations.

1.2 Literature Survey

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 et al 82], [Tsai and Huang 84], [Prazdny 79], [Roach and Aggarwal 80], [Webb and Aggarwal 81], [Fang and Huang 84a,b], [Buxton et al 84], [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 motion and structure approach is to be judged a success.

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] and [Waxman and Ullman 83,85]). Waxman and Wohn [84,85] propose that these derivatives be found by their Velocity Functional Method which consists of 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 Taylor series expansion of the x and y components of image velocity. More recently, [Subbarao 86] and [Waxman et al 87] have proposed closed form solutions for motion and structure. These algorithms basically involve solving a cubic equation and a set of decoupled nonlinear equations.

Only recently, have researchers begun to address the use of temporal information, such as temporal derivatives, in reconstruction ([Subbarao 86], [Bandyopadhyay and Aloimonos 85]). We note that others' use of temporal derivative information and our use of time-varying image velocity information differ in one important respect: we measure our information, including spatial and temporal derivative information at many different times rather than at one just one time.

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 83,85], [Buxton et al 84], and [Subbarao 86] have begun to consider the inherent sensitivity of their algorithms to random noise in the input. See [Barron 84,88] for a more detailed survey of motion and structure algorithms and their problems.

1.3 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) The use of projective transformations requires the assumption of **rigidity**; objects are assumed to be rigid. This 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. There is much psychological evidence, for example, [Gibson 57] and [Ullman 79] to support the premise that humans have a tendency to use the rigidity assumption to analyze world scenes.

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

(e) The observer's type of motion is known. The **known motion type** assumption is necessary for relating $\vec{u}$ values on a planar surface over time. We consider two types of observer motion in this paper, although we emphasize that our treatment can be generalized to other

* Also, Canadian Institute for Advanced Research

CH2664-1/88/0000/0651$01.00 © 1988 IEEE

types of observer motion as well. **Type 1** motion involves the observer translating with constant speed while he rotates his line of sight with constant angular speed. **Type 2** motion involves the observer moving with constant translational and angular speed. Given these motion types it is possible to derive equations relating both full and normal image velocity and their derivatives at one image location and time to the motion and structure parameters at another point and time.

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 measured 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], have also advocated a correspondence-less approach. The computation of image velocity may require solving the correspondence problem although there is a group of techniques based on the relationship between spatial and temporal grayvalue distributions, for example, [Horn and Schunck 81] and [Heeger 87], for determining image velocity without the need to compute correspondence.

The main concern of this paper is how accurate must the time-varying image velocity input be? Some preliminary results were presented in [Barron **et al** 87b]. Since the noise sensitivity of motion and structure is inherent in the actual nature of the problem and not in the type of solution technique used we believe other motion and structure algorithms concerned with passive navigation will exhibit noise sensitivity behaviour compatible with our results.

2 The Algorithm and Solution Technique

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. Newton's method is used to solve the equations. Hence, we need to compute a Jacobian matrix, J. Furthermore, an initial guess to the actual solution is needed to start the convergence calculation. Because we are solving nonlinear system of equations we need to be concerned about the uniqueness of any solution obtained. It has been shown that at one time there are at most 2 solutions (for planar surfaces) [Hay 66] and that they are **dual** [Waxman and Ullman 83,85] (2). Furthermore, [Hay 66] and [Subbarao and Waxman 85] have shown uniqueness over time.

3 The Sensitivity Analysis

It is well known in the Computer Vision community that motion and structure is very sensitive to noisy input. The problem is further compounded by the fact that image velocity may not be entirely due to the 3-D velocity field [Verri and Poggio 87] as variation in the image brightness patterns can be due to other phenomena as well, such as changes in lighting conditions and specular points. As a results of these problems, some authors, for example [Thompson and Kearney 86] and [Verri and Poggio 87] argue against the quantitative use of image velocity information. In this paper the discrepancy between the measured image motion and the image velocity field is considered as another source of input error.

Recently, [Subbarao 86] and [Waxman **et al** 87], have presented closed form solutions to the motion and structure (at one time) even if the surface is nonplanar. As Subbarao [86] notes, any such algorithm's error behaviour can be predicted analytically. Hence, a numerical analysis is not needed as it is redundant. However, since we are solving motion and structure using spatio-temporal image velocity data the equations we have to solve are highly nonlinear and it is doubtful that they can be decoupled. Hence, an analytic analysis seems intractable so we instead conduct a numerical analysis.

Previous error analysis has consisted almost entirely of running particular motion and structure algorithms with random noise in the input, for example [Waxman and Ullman 83,85]. In this paper we present **average** and **worst** case error results for both minimal and least squares input. Rather than compute the percent output error given X% input error we compute error amplification factors: given a certain size input error what is the size of the output error? If $\vec{\Delta f}$ is the input error and $\vec{\Delta s}$ is the output error than the L_2 error amplification factor is simply

$$\rho = \frac{||\vec{\Delta s}||_2}{||\vec{\Delta f}||_2}. \tag{3-1}$$

If $Y=\frac{||f||_2}{||\vec{s}||_2}$, where $\vec{f}$ is the input vector and $\vec{s}$ is the output vector (the solution) then X% input error causes ρY% output error.

3.1 What Constitutes a Feasible Motion and Structure Computation?

At this point we state the criteria by which we judge whether a motion and structure computation is viable. We believe a feasible motion and structure algorithm should consistently produce worst case error amplification factors of 3 or less. For $Y \approx 1$, 10% worst case error in $\vec{f}$ would then produce at most 30% error in $\vec{s}$ for a ρ value of 3. We have chosen 10% as the upper bound for measurement error; we believe image velocity data can be measured to within this accuracy. [Heeger 87] presented some encouraging results in this regard for the measurement of image velocities. Still 30% output error is not that useful, except to give a general idea about the motion. For an autonomous vehicle we probably need $\rho \leq 1$ if the computed motion and structure parameters are useful. As well, we desire a robust calculation: the condition number of the Jacobian, J, should be small, in the thousands or less, at most. Lastly, we note that the use of worst case input in our investigation is important: if we can show that motion and structure is feasible for worst case input error then it is also feasible for all types of other error. Worst case error will seldom occur in the real world.

3.2 SVD Analysis: Computing Average and Worst Case Input Error

Given a Jacobian $J_{m \times n}$, $m \geq n$ we can compute its Singular Value Decomposition (SVD) as

$$J_{m \times n} = U^T{}_{m \times n} D_{n \times n} V_{n \times n} \tag{3.2-1}$$

where U and V are unitary matrices and D is a diagonal matrix. Given the SVD of J we can compute **worst** case SVD error amplification factors as simply the inverse of the largest diagonal element of D. If d_i is that element then $U^T \hat{e}_i$ gives the worst case error direction. We note that this is actually the worst case error direction for the linear system of equations $J\vec{h}=\vec{f}$ at one iteration (the last using the correct $\vec{s}$ values) and is only an approximation for the nonlinear system in general: the smaller the input error, the smaller the nonlinear effects and the better the approximation.

We can also conduct **average** case error analysis if we make enough random error nonlinear simulations. We can compute average case SVD error amplification factors as

$$\frac{\sum_{i=1}^{1000} ||J^+ \hat{n}_i||_2}{1000} \tag{3.2-2}$$

where J^+ is the pseudoinverse of J and $\hat{n}_i$ is a random unit vector. The largest random error amplification factor should approach worst case error amplification results.

Even though worst case SVD error is only an approximation to the actual nonlinear worst case error of a nonlinear system of equations it is still much worse than random (average) case error for most motion and structure setups. As well, SVD analysis of J allows us to **predict average** and worst case performance for a nonlinear system

[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.

[2] Not only do the two sets of motion and structure parameters specify the same flow field but the **duality** equations (see [Barron **et al** 87a], [Barron 88]) allow either one to be specified analytically in terms of the other.

of equations without performing nonlinear simulation.

We can analyze the feasibility of motion and structure using full image velocities, normal image velocities, full Taylor series coefficients and normal Taylor series coefficients; we compute the appropriate J for each type of input. We can also estimate Taylor series coefficients from image velocities. SVD analysis of the corresponding J allows us to compare error amplification behaviour with the J computed for measured Taylor series coefficients.

In order to compare worst case predicted SVD error amplification factors with nonlinear simulation results we compute the worst case error direction as described above and then add X% relative worst case error to the input by scaling the worst case error so that the image velocity or Taylor series coefficient pair with the largest ratio of error size to the magnitude of the pair has X% error; hence all other pairs have $\leq X\,\%$ error in them. In the case of normal image velocities and Taylor series coefficients we compute $X\,\%$ relative error by scaling the error vector so that the largest ratio of error to normal magnitude is X%.

3.3 Computing Initial Guess Error

We add random error to the correct solution $\vec{s}$ to obtain the initial guess $\vec{s}_0$ required to start the convergence calculation in nonlinear simulation. We compute $X\,\%$ initial guess error in $\vec{s}$ by simply adding $X\,\%$ L_2 random error for each of $\vec{u}$, $\vec{\alpha}$ and $\vec{\omega}$ separately. This means that if one set of parameters is much larger than another, the latter's initial guess error is not dominated by the larger set of parameters.

3.4 Experimental Motions and Surfaces

In choosing what motion and structure parameters we should investigate we were guided by two principles: they should be realistic given our assumptions or they should have been previously reported in other work.

Table 3.4-1 shows the parameters for eight realistic motions (and their duals) that we have devised. We show the $\vec{\alpha}$ values as unnormalized to keep them as simple integers. When $\vec{\omega} \neq (0,0,0)$ we can analyze the motions for both type 1 and type 2 observer motion; hence motions 5 and 7 are type 1 motion while motions 6 and 8 are type 2 motion. The solution point $\vec{Y}_s=(20,20)$[3] is the image point where $\vec{s}$ is computed. If $\vec{Y}_s$ were changed from (20,20) to (0,0) then motions D3, D4, D7 and D8 are singular as $\vec{u}=(0,0,0)$. Motions 1-4 are analyzed assuming pure translation (i.e. $\vec{\omega}=(0,0,0)$ is known). We can analyze motions 2-4 assuming general motion as well. However, motion 1 is singular when general motion is assumed regardless of the times the inputs are measured at. Motions 5 and 6 are also singular if all the input is obtained at time 0 but can be analyzed when the input is measured over time.

The second set of motion and structure parameters we consider are Waxman and Ullman's planar motions [83,85]. Table 3.4-2 shows the 5 motions (labelled W6 to W10 to correspond to examples 6 to 10 in [Waxman and Ullman 83]) and their duals, DW6 to DW10. These are not realistic everyday motions one might expect an autonomous vehicle to undergo. We analyze these motions at time 0 only and use $\vec{Y}_s=(0,0)$, as Waxman and Ullman did. Note that motions DW7 and DW9 are singular as $\vec{u}_D=(0,0,0)$. For all the motions described in Tables 3.4-1 and 3.4-2, X_3 is 2000.

We examine motion and structure for various environmental setups. The quantities that describe the environmental setup include $\vec{Y}_s$, the solution point, the **spatial extent** and the **temporal extent**. We compute the spatial extent of a set of image points where the input is measured at as the diagonal angle of the smallest rectangle containing all the image points. We compute the temporal extent, 0–t, of a set of image points where the inputs are measured at as the times 0, $t/3$, $2t/3$ and t for full image velocity or Taylor series coefficient input and as the times 0, $t/7$, $2t/7$, $3t/7$, $4t/7$, $5t/7$, $6t/7$ and t for normal image velocity or Taylor series coefficient input.

Table 3.4-1 Original and Dual Solutions When $\vec{Y}_s$=(20,20)			
Motion	$\vec{s}$	$\vec{s}_D$	Motion
1	$\vec{u}$=(0,0,0.496976) $\vec{\alpha}$=(0,0,1) $\vec{\omega}$=(0,0,0)	$\vec{u}_D$=(0,0,0.496976) $\vec{\alpha}_D$=(0,0,1) $\vec{\omega}_D$=(0,0,0)	D1
2	$\vec{u}$=(0,0,0.496976) $\vec{\alpha}$=(1,0,1) $\vec{\omega}$=(0,0,0)	$\vec{u}_D$=(0.460963,0,0.460963) $\vec{\alpha}_D$=(0,0,1) $\vec{\omega}_D$=(0,-0.463768,0)	D2
3	$\vec{u}$=(0,0.496976,0) $\vec{\alpha}$=(0,0,1) $\vec{\omega}$=(0,0,0)	$\vec{u}_D$=(0,0,0.038826) $\vec{\alpha}_D$=(0,1,0) $\vec{\omega}_D$=(-0.5,0,0)	D3
4	$\vec{u}$=(0,0.496976,0) $\vec{\alpha}$=(1,0,1) $\vec{\omega}$=(0,0,0)	$\vec{u}_D$=(0.036013,0,0.036013) $\vec{\alpha}_D$=(0,1,0) $\vec{\omega}_D$=(-0.463768,0,0.463768)	D3
5,6	$\vec{u}$=(0,0,0.496976) $\vec{\alpha}$=(0,0,1) $\vec{\omega}$=(0,0.2,0)	$\vec{u}_D$=(0,0,0.496976) $\vec{\alpha}_D$=(0,0,1) $\vec{\omega}_D$=(0,0.2,0)	D5,D6
7,8	$\vec{u}$=(0,0.496976.0) $\vec{\alpha}$=(1,0,1) $\vec{\omega}$=(0.2,0,0)	$\vec{u}_D$=(0.036013,0,0.036013) $\vec{\alpha}_D$=(0,1,0) $\vec{\omega}_D$=(-0.263768,0,0.463768)	D7,D8

Table 3.4-2 Waxman and Ullman's Planar Motions and Their Duals At $\vec{Y}_s$=(0,0)			
Motion	$\vec{s}$	$\vec{s}_D$	Motion
W6	$\vec{u}$=(6,5,4) $\vec{\alpha}$=(-1,0,-1) $\vec{\omega}$=(3,2,1)	$\vec{u}_D$=(4,0,4) $\vec{\alpha}_D$=(-1.5,-1.25,-1) $\vec{\omega}_D$=(-2,4,6)	DW6
W7	$\vec{u}$=(6,5,0) $\vec{\alpha}$=(-1,0,-1) $\vec{\omega}$=(3,2,1)	$\vec{u}_D$=(0,0,0) $\vec{\alpha}_D$=(1.2,1,0) $\vec{\omega}_D$=(-2,8,6)	DW7
W8	$\vec{u}$=(1,2,3) $\vec{\alpha}$=(-1,-2,-1) $\vec{\omega}$=(4,5,6)	$\vec{u}_D$=(3,6,3) $\vec{\alpha}_D$=(-0.333,-0.667,-1) $\vec{\omega}_D$=(8,3,6)	DW8
W9	$\vec{u}$=(1,2,0) $\vec{\alpha}$=(-1,-2,-1) $\vec{\omega}$=(4,5,6)	$\vec{u}_D$=(0,0,0) $\vec{\alpha}_D$=(1,2,0) $\vec{\omega}_D$=(2,6,6)	DW9
W10	$\vec{u}$=(6,5,4) $\vec{\alpha}$=(0,0,-1) $\vec{\omega}$=(3,2,1)	$\vec{u}_D$=(0,0,4) $\vec{\alpha}_D$=(-1.5,-1.25,-1) $\vec{\omega}_D$=(-2,8,1)	DW10

4 Experimental Results

We present qualitative rather than quantitative results here; complete numerical results are in [Barron 88]. We consider both **minimal** (the least amount of data needed to recover motion and structure) and **least squares** input.

4.1 Using Minimal Input

In this section we investigate the use of minimal full/normal image velocity or Taylor series coefficients data in the computation of motion and structure. The two main questions we are concerned with are "How sensitive is the computation to noisy input?" and "Can anything be done to reduce the noise sensitivity?".

4.1.1 Using Minimal Image Velocity Input

Since four image velocities constitute minimal input for general motion we measure each image velocity at the four corners of a square centered at $\vec{Y}_s$ at each of the four times specified by the temporal extent [4]. For each motion, we use a temporal extent 0–t where t is varied from 0 to 1 for motions 1-4, 7 and 8 and 0.3 to 1 for motions 5 and 6. For motions W6-DW10 t is fixed at 0. If $\vec{Y}_s$=(20,20) the spatial extent is 30.5^o while if it is (0,0) the spatial extent is 30.9^o.

[3] Given f=1 and that the minimum/maximum horizontal/vertical coordinates of the image are $\pm\frac{1}{2}f$ if we assume a 256×256 pixel image coordinate system $\vec{Y}_s$=(20,20) in pixels corresponds to $\vec{Y}_s$=(0.078125,0.078125,1) in f units. We use pixel image locations in this paper because these can be expressed as simple integers and are more intuitive.

[4] If pure translation is assumed we only need 2½ image velocities. In this case we let the LU decomposition of J choose the 2½ image velocities to be used in the motion and structure computation.

Average and worst case SVD results show that the computation of motion and structure is untenable. Indeed, even if the feasibility criterion of a maximum worst case error amplification factor of 3 were increased to 10 most of the motions are still not viable. Worst case error amplification factors ranged from a low of 7 to a high of 317 for motions 1-8 and from 3.4 to 36 for motions W6-DW10. Some of the motions, especially motion 5 had average and worst case error amplification factors in the 100's! However, we are able to make two important observations:

(1) The worst case error amplification factors observed from the SVD analysis and the corresponding nonlinear simulations showed good agreement: the SVD amplification factors are good predictors for nonlinear simulation behaviour. We used 0% to 1.4% relative worst case image velocity error (because of the high error amplification factors, larger input error usually did not allow convergence) and 0% initial guess error in the nonlinear simulations.

(2) In general, worst case error amplification factors were at least 2-3 times higher than average case error amplification factors. However, the maximum of the random error amplification factors used in the computation of an average case error amplification factor for a particular motion was quite close to the worst case error amplification factor for that motion. Worst case error often occurred for minimal input. It seems that worst case analysis should be an important feature of any sensitivity analysis.

The second question we posed, "Can anything be done to reduce the noise sensitivity?" can be answered more positively. In particular, we demonstrate that time helps. As the temporal extent is increased, the average and worst case error amplification factors decrease for most of the motions. As well, increasing the spatial extent helps for all motions. Indeed, for some of the motions a spatial extent of 70^o (the full image plane) yield error amplification factors less than 3 (or, more commonly, in the 4-6 range).

We are able to report numerous other results:

(1) Worst case error amplification factors for full image velocities are only slightly larger than worst case difference error amplification factors where error is added to the image velocity differences only: worst case error is almost entirely due to error in the image velocity differences. This means large image velocity mean error can be handled with ease. Some types of autonomous vehicle motion, for example, hitting a rut in a road, introduce this type of error into the image velocity input.

(2) It is reasonable to question the validity of our algorithm's underlying assumptions. As Thompson and Kearney [86, p17] comment:

> "Unrealistic assumptions are only justifiable when it can
> be shown that useful answers can be obtained in realistic
> situations despite violations in the assumptions".

We test violation of the planarity, rigidity and same surface assumptions by perturbing $\vec{\alpha}$ values in one of three ways: (a) by measuring the image velocities on two halves of a hinged surface, (b) by .measuring the image velocities on a spherical surface and (c) by perturbing the $\vec{\alpha}$ values randomly. We test violation of the fixed axis assumption by adding randomly generated error to $\vec{\omega}$ at each time an image velocity is measured at while we test violation of the constant observer translation assumptions by adding random error to $\vec{u}$ and X_3 at each time an image velocity is measured. We perform nonlinear simulation, using assumption error scaled to be the same size as worst case image velocity error. The results show that the assumption error amplification factors are usually significantly less than worst case error amplification factors. It seems that we should be more concerned with image velocity accuracy rather than with violation of the algorithm's underlying assumptions.

(3) We consider several situations where computing a restricted set of motion and structure may be realistic. One, in particular, is if an autonomous vehicle's environment is restricted to the corridors of a building. In this case, u_2, ω_1 and ω_3 are known to be zero. Hence, there are only five independent motion and structure parameters in the general case and only four when pure translation is assumed. Of course, the planarity assumption is satisfied for most of the vehicle's environment as it is primarily composed of walls, floors and ceilings. We examined motions 1-2 and 5-6 under the assumption of planar translation, vertical rotation. Using the minimal amount of image velocity data needed yields worst case error amplification factors (from 16 to 220) that are actually larger than those observed in the initial sensitivity for the corresponding general motions. However, using all four image velocities resulted in lower worst case error amplification factors than for the corresponding general motions. In this and other cases (see [Barron 88]), knowing something about the motion made the motion and structure calculation more viable (although not necessarily feasible).

(4) The last set of results that we report for minimal image velocity input is the effect of the initial guess on the output. In particular we demonstrate the following properties of our algorithm:

(a) For 0% image velocity error we are able to solve motion and structure for large initial guess error, up to 100% and more. More interestingly, we observed the presence of multiple (nondual) solutions for motions 5-7. These multiple solutions produced flow fields distinct from those for the original solution but both sets of flow fields shared four common image velocities at the image locations and times where the input was measured at. Examples of multiple solutions are reported in [Barron et al 87a] and [Barron 88].

(b) We also observed the presence of dual solutions for motions W6, W8, W10 and 2. No multiple solutions were observed for these motions; it is unlikely that two distinct nonsingular flow fields would have four common image velocities. Again 0% image velocity error and up to 100% initial guess error was used in the nonlinear simulations.

(c) Finally, we performed nonlinear simulation, varying worst case image velocity error from 0% to 1.4% for fixed initial guess errors of 50% and 100%. For most solved runs we observed that image velocity and initial guess error are independent of each other. The other solved runs produced multiple solutions in the presence of nonzero image velocity error; these solutions are due to the initial guess error as discussed above in (a).

4.1.2 Using Minimal Normal Image Velocity Input

The recovery of motion and structure from eight normal image velocities measured at eight times is also possible (or from five normal image velocities measured at five times if pure translation is assumed) provided that the flow field exhibits enought structure. Waxman and Ullman [83,85] call this the "aperture problem in the large" versus the usual "aperture problem in the small" [Marr and Ullman 81]. We measure eight normal image velocities, $\vec{v}_n$, at the four corners of a square and the four midpoints of the square's sides. The normal image velocity data is generated in one of two ways: either by alternately using the horizontal and vertical components of the full image velocity at each of the eight image points and times or by computing eight unit random vectors, $\hat{n}$ and then computing $\vec{v}_n=(\vec{v}\cdot\hat{n})\hat{n}$ for the image velocity values, $\vec{v}$, measured at each of the eight image points and times.

Again, we perform SVD analysis and nonlinear simulation for all 16 motions and find good correlation between predicted and observed worst case error amplification factors. Typically, normal image velocity error amplification factors are about twice as large as the corresponding error amplification factors for full image velocity input when the same spatio-temporal extent is used.

As in the previous section we also observed multiple solutions (two distinct sets of eight normal image velocity flow fields shared eight common normal image velocities) for motions 1-2 and 5-8. Dual solutions were found for motions 7-8 and W6, W8 and W10 when the temporal extent is 0-0. Full details are in [Barron 88].

4.1.3 Using Minimal Taylor Series Coefficient Input

In this section we report results obtained using minimal Taylor series coefficient input. We can use either

$$\vec{g}_s=\left[\vec{v},\frac{\partial\vec{v}}{\partial y_1},\frac{\partial\vec{v}}{\partial y_2},\frac{\partial^2\vec{v}}{\partial y_1\partial y_2}\right], \tag{4.1.3-1}$$

or

$$\vec{g}_t=\left[\vec{v},\frac{\partial\vec{v}}{\partial y_1},\frac{\partial\vec{v}}{\partial y_2},\frac{\partial\vec{v}}{\partial t}\right] \tag{4.1.3-2}$$

to compute motion and structure. At any single time, $\vec{g}_s$ fully specifies the flow field (see [Waxman and Ullman 83,85] and [Barron 88]). $\vec{g}_s$ coefficients estimated from a single image velocity field are identical to those measured from the image intensity data at the same time. As well, we note that we can measure the $\vec{g}_s$ coefficients in time and we can measure $\vec{g}_t$ at one time if we wish.

The first observation we make is that small image velocity error corresponds to much larger Taylor series coefficient error. This means that the error amplification factors resulting from the use of Taylor series coefficient input will be considerably smaller than from the use of image velocity input.

Nonlinear simulation results for estimated and measured Taylor series coefficients agree reasonably well with predicted SVD error amplification factors. Unfortunately, except for isolated cases, the computation of motion and structure is still not feasible for minimal Taylor series coefficient input. These isolated cases consist of about half the runs for motions 1-2, all runs for motions 3-4, about a quarter of the runs for motions 7-8 and a few of Waxman and Ullman's planar motions. These solved cases usually occurred for a spatial extent of 0^o and a temporal extent of 0–t, $0\le t\le 1$, when $\frac{\partial^2\vec{v}}{\partial y_1\partial y_2}$ was used and for large spatio-temporal extents when $\frac{\partial\vec{v}}{\partial t}$ was used. Even though minimal Taylor series coefficient input is still not feasible it is much more encouraging than minimal image velocity input.

We define estimation error as the difference between estimated and actual Taylor series coefficients for a particular spatio-temporal extent (the spatial extent is always nonzero). For larger temporal extents this error can be quite large as the estimated and measured $\frac{\partial\vec{v}}{\partial t}$ differ significantly. Indeed, we obtain nonlinear convergence for estimation errors of up to 40%. The error amplification factors were about 1-2 but could be as high as 5. Of course, estimation error is not worst case error.

A comparison of the worst case error amplification factors that result when $\frac{\partial\vec{v}}{\partial t}$ or $\frac{\partial^2\vec{v}}{\partial y_1\partial y_2}$ is used show that generally $\frac{\partial^2\vec{v}}{\partial y_1\partial y_2}$ is better than $\frac{\partial\vec{v}}{\partial t}$. However, for these motions $\frac{\partial\vec{v}}{\partial t}$ was usually either much larger or much smaller than the other $\vec{v}$, $\frac{\partial\vec{v}}{\partial y_1}$ and $\frac{\partial\vec{v}}{\partial y_2}$ coefficients. The one exception is motion 5 where $\frac{\partial^2\vec{v}}{\partial y_1\partial y_2}$ is much smaller that $\frac{\partial\vec{v}}{\partial t}$ and, as a result of being closer to the size of the other coefficients, produces slightly smaller worst case error amplification factors than $\frac{\partial\vec{v}}{\partial t}$.

In the case of measured Taylor series coefficients we can collect all data at one image point and time (as suggested by [Longuet-Higgins and Prazdny 81] or [Waxman and Ullman 83,85]) or at four image points and times. Typically, the use of $\vec{g}_s$ at one point and time produced the best results. Using $\vec{g}_t$ measured for small spatio-temporal extents yields the largest error amplification factors but these are significantly reduced for larger spatio-temporal extents. $\vec{g}_s$ can also be used for nonzero spatio-temporal extents, producing much better results than observed from the use of $\vec{g}_t$. Lastly, we did not observe any multiple solutions even when 100% initial guess error was used, although all runs solved.

4.1.4 Using Minimal Normal Taylor Series Coefficient Input

We measure normal (horizontal/vertical and random) Taylor series coefficients at the same times and image point offsets as in the normal image velocity case. A set of normal Taylor series coefficients consists of two normal coefficients for each of $\vec{v}_n$, $\frac{\partial\vec{v}_n}{\partial y_1}$, $\frac{\partial\vec{v}_n}{\partial y_2}$ and $\frac{\partial\vec{v}_n}{\partial t}$ or $\frac{\partial^2\vec{v}_n}{\partial y_1\partial y_2}$. As before, there is a good correlation between predicted and observed worst case error amplification factors for minimal input data. Error amplification factors range from 1.7 to 552 for random normal Taylor series coefficient data and from 5 to 780 for horizontal/vertical normal Taylor series coefficient data. We also observed the presence of multiple solutions for motions 1, 5, 6 and 7 and dual solutions for some of Waxman and Ullman's planar motions.

4.2 Using Least Squares Input

As we have seem above, motion and structure is not feasible for minimal input. In this section we report the results of our investigation when least squares input is used. The relevant question here is "How much least squares input is needed to obtain a feasible motion and structure calculation?". Many researchers, for example Waxman and Ullman [83,85] have suggested that the use of least squares input can produce acceptable error amplification factors. A second related question is "With what kind of input and how much of it is needed?".

The main tool used in our analysis of least squares input is SVD analysis. We compute average and worst case error amplification factors for various amounts of least squares input for full and normal image velocity and full and normal Taylor series coefficients. We construct least squares input by using multiple sets of minimal input data in our experiments. One advantage of using time-varying input is that it is easier to collect the volume of data needed for a feasible motion and structure calculation over a period on time rather than at one time only. Lastly, we emphasize that all the experimental results presented in section 4.1 for minimal input also hold for least squares input as well. Hence, increasing the temporal extent for least squares image velocity data will result in lower error amplification factors for most of the 16 motions.

4.2.1 Using Least Squares Image Velocity Input

In this and the next section we use a solution point $\vec{Y}_s=(0,0)$. We investigate least squares using 3, 6 and 24 sets of image velocity data. (Each set of image velocity consists of four image velocities.) The image points are distributed within the same spatio-temporal extents as used for minimal input (see [Barron 88] for details). 24 sets of image velocity means we use 96 image velocities in total (we compute $J_{192\times 8}$). We believe this is an upper limit on the number of image velocities we can realistically hope to measure for a single surface patch over a small range of times.

While the use of least squares image velocity data is helpful in reducing worst case error amplification factors it still does not make motion and structure computations feasible in general, although motions 3, 4 and DW10 which are now feasible. It would seem that the tradeoff of obtaining a worst case error amplification factor of 3 or less and the necessary increase in the amount of data and the spatial and temporal extents of the measurements required is unreasonable. Using more than 100 image velocities, a larger spatial extent than 30.9 degrees (already about 1/5 of the image) or a larger temporal extent (the underlying assumptions such as local planarity and the fixed axis assumption may be violated) is not realistic for most motions. On the other hand, average SVD amplification results for 24 sets of image velocities are most encouraging; all motions except motion 5 can be solved (it is solved when the temporal extent exceeds 0-0.5). In fact, except for motion 5 average case error amplification factors ranged from 0.1 to 1. If we examine the eight error amplification factors that result from the SVD of J we see that only one (or sometimes two) of the factors are large, the others are often around 1 or less. It seems that worst case error directions occupy only a small part of the overall error direction space. This suggests that average error analysis rather than worst case error analysis may be a more appropriate type of analysis for least squares image velocity input. This is an important observation: all researchers, including [Waxman and Ullman 83,85], [Waxman and Wohn 84,85], [Subbarao 86], [Bandyopadhyay 86] and others, used random noise in the input to their motion and structure algorithms.

4.2.2 Using Least Squares Normal Image Velocity Input

As in the minimal input case we use two types of normal image velocity data: horizontal/vertical and random. One reason to use normal image velocities directly in the computation of motion and structure is because it is easier to measure normal rather than full image velocity data from most image sequences (due to the aperture problem [Marr and Ullman 81]). Another of our goals in using normal image velocity data was to test the Velocity Functional Method [Waxman and Wohn 84,85]. The Velocity Functional Method was introduced as a way of computing the "observables" from a normal image velocity field. These observables were then used in the subsequent motion and structure calculation. We carried the investigation one step further: we are interested in computing motion and structure directly from least squares normal image velocity input. This combines the Velocity Functional Method and the motion and structure computation from the "observables" into a single computation. We found that even when 96 normal image velocities were used the motion and structure computation was only feasible for motions 3, 4 and DW10. We obtained worst case error amplifications of about 10-20 for Waxman and Ullman's motions (with the exception of motion DW10 which exhibited an error amplification less than 2) when 96 normal image velocities were used. However when average SVD analysis was performed the error amplification ranged from 0.1 to 2 or less for all motions, except motion 5 (where the average error amplification ranged from just over 2 to just under 14). Waxman and Wohn report 5% output error given 10% random input error for typical runs. (Depending on the L_2 of the input and output vectors this is equivalent to an error amplification factor ranging from 0.5 and 2.0.) Hence, their results agree quite well with our predictions.

4.2.3 Using Least Squares Taylor Series Coefficient Input

The use of worst case least squares Taylor series coefficient data allows motion and structure to be solved for all the motions except motion 5. We use two point formulations when computing motion and structure from actual Taylor series coefficients: the first is the same as for the least squares image velocity data and the second measures each set of Taylor series coefficients at a particular image point and time. The same spatio-temporal extents are used for both point formulations. Estimated Taylor series coefficients are computed using image velocities measured at the first point formulation. Complete details are in [Barron 88]. The error amplification factors for the two point formulations are very similar. All motions, with the exception of motion 5, are feasible using only 6 sets of Taylor series coefficients, a reasonable amount of data to measure, especially in time. Motion 5 can be solved if average case error is present in the input. It seems that actual Taylor series coefficients are a bit better than estimated Taylor series coefficients: the worst case error amplifications are slightly smaller.and various actual Taylor series coefficient pairs can be measured at a large number of spatial-temporal extents or complete sets can be measured at individual image locations and times. At the beginning of Section 4.1.3 we described the relationship between error in a minimal image velocity set and its corresponding estimated Taylor series coefficients. Small image velocity error corresponds to much larger Taylor series coefficient error. Hence, it is not surprising that Taylor series coefficients are more robust than image velocities for motion and structure computations when there is noise in the input. The problem of image velocity computation has been addressed by many researchers (see Section 1 for references) without overwhelming success. We suggest that a better approach might be to measure actual Taylor series coefficients directly from the raw time-varying data. The design of spatio-temporal filters to perform this task seems appropriate.

4.2.4 Using Least Squares Normal Taylor Series Coefficient Input

Again, because of the aperture problem it will be easier to measure normal Taylor series coefficients rather than full Taylor series coefficients. With the exception of motion 5, these can be solved for worst case input error provided all 96 normal Taylor series coefficients are used. This includes both horizontal/vertical and random normal Taylor coefficient series data measured for both point formulations. Motion 5 can be solved if average Taylor series coefficient error is used. Since the spatio-temporal filters that would measure this type of data would do so locally we believe that this type of data is the most practical for motion and structure calculations.

4.3 The Stability of Motion and Structure

We have observed that error amplification factors decrease for increasing spatio-temporal extents. In addition, as the amount of input data is increased not only is the error amplification reduced but the condition number also decreases. This means that least squares motion and structure computations become both more feasible and more robust than the same calculation from a minimal set of input data. Thus, our feasibility criteria is more likely to be satisfied with larger amounts of input data.

5 Conclusions

The main contributions of this research are:

(1) We have formulated a motion and structure algorithm that use full/normal image velocity data and full/normal Taylor series coefficient data as its input. Measuring this input over time usually results in lower sensitivity to input error, allows us to analyze some motions that are singular at one time and allows us to collect larger amounts of input data for a least squares calculation.

(2) We have conducted an extensive sensitivity analysis for both minimal and least squares input for the four types of input for average and worst case error. We found that:

(a) Motion and structure can be feasibly recovered in the average case if full/normal least squares image velocity data is used.

(b) Motion and structure can be feasibly recovered in the worst case if full/normal least squares Taylor series coefficient data is used.

Indeed, since most of the motions exhibit ρ values less than 1 for these average or worst case results we conclude that the motion and structure paradigm is suitable for autonomous vehicle navigation. This is in sharp contrast to the popular belief in the Computer Vision community that motion and structure is not suitable for such an application because of its sensitivity to noisy input. Further, we recommend the following: use normal Taylor series coefficient data measured over time to compute motion and structure. The design of filters to measure this normal data locally from an image sequence is advocated as an area of future research. We have also devised a binocular motion and structure algorithm [Barron et al 87c] that uses time-varying image velocity from a binocular image sequence but have yet to perform a sensitivity analysis.

Bibliography

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

Bandyopadhyay A. and J. Aloimonos, "Perception of Rigid Motion from Spatio-Temporal Derivatives of Optical Flow", TR-157, Dept. of Computer Science, University of Rochester, N.Y., March 1985.

Bandyopadhyay A. "A Computational Study of Rigid Motion Perception", PhD Thesis, TR-211, Dept. of Computer Science, University of Rochester, Dec. 1986.

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

Barron, J.L., "Computing Motion and Structure From Time-Varying Image Velocity Information", PhD thesis, June, 1988, Dept. of Computer Science, University of Toronto. (also RBCV-TR-88-24, Dept. of Computer Science, University of Toronto, August, 1988.)

Barron J.L., A.D. Jepson and J.K. Tsotsos, "Determination of Egomotion and Environmental Layout From Noisy Time-Varying Image Velocity in Monocular Image Sequences", 4^{th} Intl. Conference on Image Analysis and Processing, Sicily, Italy, Sept. 1987.

Barron J.L., A.D. Jepson and J.K. Tsotsos, "The Sensitivity of Motion and Structure Computations", AAAI, 1987, pp700-705.

Barron J.L., A.D. Jepson and J.K. Tsotsos, "Determining Egomotion and Environmental Layout From Noisy Time-Varying Image Velocity in Binocular Image Sequences", IJCAI, August, 1987, pp822-825.

Bruss A.R. and Horn B.K.P., "Passive Navigation", CVGIP (formally CGIP), Vol. 21, 1983, pp3-20.

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

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

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

Fang J.-Q. and Huang T.S., "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.

Gibson J.J., "Optical Motions and Transformations as Stimuli for Visual Perception", Psychological Review, Vol. 64, No. 5, 1957, pp288-295.

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

Heeger D.J., "Optical Flow from Spatiotemporal Filters", ICCV, 1987, pp181-190.

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

Kanatani K, "Structure from Motion without Correspondence: General Principle", IJCAI, 1985, 886-888.

Lawton D.T., "Processing Translational Motion Sequences", CGIP22, pp116-144.

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

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

Marr D. and Ullman S., "Directional Selectivity and its Use in Early Visual Processing", Proc. Royal Society of London, B211, 1981, pp151-180.

Prazdny K., "Motion and Structure From Optical Flow", IJCAI, 1979, pp702-704.

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

Subbarao M. and Waxman A.M., "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, 1985. (Also, 3^{rd} Workshop on Computer Vision: Representation and Control, 1985.)

Subbarao M., "Interpretation of Visual Motion: A Computational Study", CAR-TR-221, Center for Automation Research, University of Maryland, PhD Thesis, Sept., 1986.

Thompson W.B. and Kearney J.K., "Inexact Vision", Proc. Workshop on Motion: Representation and Analysis, May 7-9, 1986, pp15-21.

Tsai R.Y., Huang T.S. and Zhu W.-L., "Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch II: Singular Value Decomposition" IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. 30, No. 4, August, 1982, pp525-534.

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

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

Verri A. and Poggio T., "Against Quantitative Optical Flow", ICCV, 1987, pp171-180.

Waxman A.M. and Ullman S., "Surface Structure and 3-D Motion from Image Flow: A Kinematic Analysis", CAR-TR-24, Center for Automation Research, University of Maryland, Oct., 1983.

Waxman A.M. and Ullman S., "Surface Structure and Three-Dimensional Motion From Image Flow Kinematics", Intl. Journal of Robotics Research, Vol. 4, No. 3, Fall, 1985, pp72-94.

Waxman A.M. and Wohn K., "Contour Evolution, Neighbourhood Deformation and Global Image Flow: Planar Surfaces in Motion", CAR-TR-58, Center for Automation Research, University of Maryland, April, 1984. (Also in Intl. Journal of Robotic Vision, Vol. 4, No. 3, Fall, 1985, pp95-108.)

Waxman A.M., Kamgar-Parsi B. and Subbarao M., "Closed-Form Solutions to Image Flow Equations", ICCV, 1987, pp12-24.

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