MOTION FROM A SEQUENCE OF IMAGES

Igor Pavlin*

University of Massachusetts at Amherst

COINS Technical Report 87–132

February 1988

ABSTRACT

This paper considers the problem of the recovery of motion parameters of a rigid object moving through environment with constant but arbitrary linear and angular velocities. The method uses temporal information from a sequence of images such as those taken by a mobile robot. Spatial information contained in the images is also used. The temporal sequence, combined with the assumption of constant velocities, provides powerful constraints for the motion trajectory of rigid objects.

We derive a closed form solution for the rigid object trajectory by integrating the differential equations describing the motion of a point on the tracked object. The integrated equations are non-linear only in angular velocity and are linear in all other motion parameters. These equations allow the use of a simple least-square error minimization criterion during an iterative search for the motion parameters. Experimental results demonstrate the power of our method in fast and reliable convergence.

*Now at the Institute of Robotics and Intelligent Systems at the University of Southern California.

This work was supported by the Defense Advanced Research Agency, under grants N00014-82-K0464, and DACA 76-85-C-0008.

1 INTRODUCTION

1.1 Background

The interest in temporal analysis of image sequences has sharply risen with the increase of available computational power [10]. Bolles and Baker argue that equal weight be given to both temporal and spatial information contained in a sequence of images (frames). Two-image motion algorithms often use spatial integration [18]. However, in a two-image analysis the only temporal information that can be used is differential (positions of image points in two close time instances). This information is too local in scope and too sensitive to errors. In this paper we argue for more extensive use of temporal information. Temporal information is obtained from measurements of distinctive feature positions in the image plane in several time instances (images). A distinctive feature is labeled with both spatial and temporal coordinates and is called an image event. The time coordinate is treated on an equal footing with spatial coordinates. Thus, the idea of smoothing can be applied to time coordinate as well, resulting in reduced sensitivity to small errors in spatial coordinates. Another important feature of our method is the integration of temporal information which is achieved by solving differential equations of motion. By integrating temporal information over several images we get more reliable information about the motion trajectory.

There is evidence that temporal integration is performed by humans, as demonstrated in [21]. It is only the integration of temporal information that is discussed here in greater detail.

Computations of displacement fields [1,6,16,17,22,23,24,38,39] by their nature give more weight to the spatial information in image features. Computations are sensitive to noise in the image because they rely on derivatives of feature displacements. The motion parameters are obtained from a set of high-order non-linear equations, resulting in costly computations, low precision and unstable solutions. The deficiencies of these approaches are especially visible in problems of motion segmentation and occlusion. Motion algorithms using multiple images from an image sequence have been developed by many authors [2,7,10,14,32,34,42].

The research in this paper has been motivated by the problem formulation of Shariat [34] finding the motion parameters of a rigid object moving with constant translational and rotational velocities, under the assumption that five images are equally spaced in time. These assumptions enable Shariat to propose a set of difference equations relating the positions of features in the image plane and the motion parameters of the projected point. The solution he obtains is a set of non-linear polynomial equations in unknown motion parameters. The equations are of a rather high (5th) order. The motion parameters are obtained using a Gauss-Newton non-linear least-square method with carefully designed initial-guess schemes.

Our work also relates to that of Broida and Chellappa [13,14] due to the fact that it considers estimation of motion parameters of a rigid body from several images. Broida and Chellappa [14] are estimating kinematic parameters of a rigid body and its structure by tracking several feature points in a sequence of images. The object of their work is to estimate the influence of white noise in feature positions on the recovery of motion parameters and to find the robustness of the method to bad initial guesses. This is important because the point feature extraction methods return feature positions below the expected precision of the algorithm. Their work can be used in the prediction of the feature positions in the next frame(s). It is based on iterative Kalman filtering techniques [12], and stochastic estimation techniques. (good references on the theory behind stochastic estimation techniques used in their work can be found in [20,27]). Wünsche [41] has a similar approach. The prediction of the feature in the next frame usually leads to a reduced search space in the feature correspondence problem, but we have not considered this important issue in the present paper.

Our work concentrates on the computer vision aspect of the problem, and in a different way proves that using several frames is a good way to detect objects moving in the camera's field of view or detect camera movement. We argue, in the sense of Lowe's [25] uniqueness of the viewpoint constraint, that the image positions of the same feature in several frames (i.e., several time instances) provide a powerful constraint on the possible types of object motion. We might call the time constraint "the uniqueness of the time sequence constraint". This constraint reflects deterministic motion of a rigid body whose initial position and velocities are known. The time constraint combined with the Lowe's spatial constraint is an excellent tool for detection of rigid body kinematics and shape from motion.

Our approach begins with the set of differential equations describing the motion of a point P on a rigid body moving with constant translational and rotational velocity through the environment. The equations are solved analytically, i.e., a set of parametric equations $P = (X(t), Y(t), Z(t))$ describing the helix-like trajectory of the point P is found. The parameters in these equations are the initial position and the linear and angular velocity of the point P. Our goal is to determine these parameters from the central projection of the trajectory of the point P on the image plane.

The central projection of the point P on the image plane is labeled Q. This distinctive feature is characterized by image coordinates x_i, y_i and the time label t_i of the image: $Q_i = (x_i, y_i, t_i)$, $i = 1, \ldots, n$, where n is the number of considered images. We will call Q_i image events. Under the assumptions of the known projective geometry and constancy of motion parameters, only a few image events Q_i are needed to find the motion parameters of the environmental point P. The reason is that we know (from the closed form solution for the trajectory) what constraints exist between image events Q_i and the motion parameters of the point P. The constraints are non-linear only in rotational parameters and the type of non-linearity is known exactly. In our method, there is no constraint imposed on the time-interval between images, and the equations can be more readily adapted for motion of non-rigid objects, as well as for accelerated motions by using perturbation techniques [15,29].

We do not consider the problem of feature correspondence over several images. There are indications that the correspondence problem can be handled quite successfully under the conditions of an approximately known displacement path [3,4,5,8,9]. Bharwani uses correlation measurements and the history of the motion to predict the position of a point feature in the next frame using ths concepts developed by Anandan. This approach has some difficulties in trying to predict the feature position with accuracy less than a half of pixel, due to the shallow nature of the correlation measure at this resolution. The feature prediction was proven to be very helpful both for efficiency and for more reliable feature matching.

Sethi and Jain SET87, for example, use "smoothness of motion" to reduce the search explosion of possible correspondences of several image features through several images. Once a good initial guess for motion parameters is found the search space of correspondences can be drastically reduced and possible correspondences become better defined as the computation proceeds.

We plan to attack the feature correspondence problem by using symbolic features such as lines, and perhaps regions. In the case of line matching we can choose feature points to be at the intersection of two lines, and the line correspondence can be established using methods similar to those in Medioni [28]. In this case the major practical problem is to establish the correct line correspondence. The extra effort is rewarded by greater robustness to noise.

The other possibility is to choose lines themselves as features. Spetsakis and Aloimonos [36] use three frames and 13 line correspondences to establish the translational and rotational parameters of the object whose lines are being tracked. They develop a method resulting in a linear system of equations, that authors report to be noise sensitive. We plan to develop a theory that will track the motion of lines through several frames. The theory will have similar structure as the one presented in this paper and it will be a non-linear theory with an explicit frequency dependence. We expect the theory to be less noise sensitive than the one reported by [36].

More recently Williams [40] have developed a spatio-temporal grouping algorithm that produces token-based line correspondences across a sequence of images. This algorithm appears to be robust and utilizes two other algorithms developed at the University of Massachusetts, a spatial line grouping algorithms by Boldt and Weiss [11] and Anandan's algorithm for developing displacement fields with confidence measures [4].

In the case of region matching we can track region centers (like center of mass) as in work by Price [31]. Use of line and region correspondence should result in a more noise-robust approach to the correspondence problem, although we predict some difficulties with the precision of these methods.

One important issue not considered here is the choice of a good initial guessing scheme. In this paper we used simple parameter estimation techniques (such as the sign of the path curvature, direction of motion, magnitude of motion) to predict the initial velocities and position of the tracked feature. We hope to incorporate more sophisticated techniques of initial estimation similar to those seen in work of Shariat [34]. The initial guessing methods used by Lowe [25] in his iterative solution for 2-D to 3-D object recognition are also appropriate for this approach. Lowe's method is also a Newton-Ralphson technique (just like ours) and was proven to be very stable and converged to the right solution in almost all cases. We found that this is the case with our method, too. The good convergence is probably due to the viewpoint and time constraints discussed earlier. Therefore, even with initial guesses that are "far" from the correct solution, the methods converge to the correct solution.

We use a set of noise-free synthetic images to demonstrate the feasibility and speed of the approach, and its high accuracy in noise free conditions. At the time of the publication of this report, we have not obtained results for motion sequences from a natural environment. We expect that the performance will not be as good as reported here, but the use of symbolic features should greatly improve the robustness of the method.

In Section 2 we establish the relevant set of equations, in Section 3 we give a method of solution, in Section 4 we present preliminary results and in Section 5 we discuss further work.

2 Establishing the Equations

In this section, we first derive motion equations for general object motion. We then introduce the assumptions of object rigidity and constant linear and angular velocity and solve the simplified equations in closed form.

Our analysis of general motion [15] is slightly different from the analysis of a camera moving through the environment [24], since the possibility of tracking several objects is kept in consideration. Also, our method of solution is different from that of Shariat [34] and the result is more general. We integrate the differential equations of motion and derive equations that are simpler and more powerful since they contain exact and explicit information about the body motion and they are not restricted to equally-spaced time images.

The camera setup and notation are presented in Figure 1. For the purpose of clarity, we distinguish between 3 coordinate systems: a reference coordinate system with the center at point O (camera-centered coordinate system); an intermediate coordinate system, which is here a center-of-mass coordinate system with the origin C and coordinates (X_c, Y_c, Z_c); and a body-fixed coordinate system (X_b, Y_b, Z_b) with the origin also at the point C. The center-of-mass coordinate system is translating with axes parallel to the axes of the camera-centered coordinate system and the body-fixed coordinate system is rotating around C. The intermediate coordinate system is introduced in order to allow the separation of rotational motion (and perhaps the motion internal to the body-fixed coordinate system) from the rest of the motion. The choice of C is arbitrary for our purposes, and any other point is a valid choice for the center of rotation. $0xy$ is the image coordinate system.

The position of an arbitrary point P on a moving object is characterized according to Figure 1 by vector $R(t)$ (in matrix notation)

$$R(t) = C(t) + r(t) \tag{1}$$

where $C(t)$ is the current position of C and $r(t)$ is the current position of the point P relative to C

in both the center-of-mass and body-fixed coordinate systems. Differentation of Eq. (1) gives [15]

$$\left(\frac{dR(t)}{dt}\right)_{camera} = \left(\frac{dC(t)}{dt}\right)_{camera} + \left(\frac{dr\,(t)}{dt}\right)_{body} + \omega \times r. \tag{2}$$

$\left(\frac{dC(t)}{dt}\right)_{camera}$ is the contribution to the speed of point P coming from the motion of the center-of-mass coordinate system and the term $\omega \times r$ defines changes in the position of the point P due to the instantaneous rotation ω of the body around C. The term $\left(\frac{dr\,(t)}{dt}\right)_{body}$ is the internal velocity of the point P in the body-fixed coordinate system. This velocity is zero if the body is rigid.

We now introduce the assumptions that the object is rigid, and has constant translational and rotational velocity. The equation of motion of the point P in the camera-centered coordinate system is then the solution of the following pair of equations, describing a constant translational motion V of the origin C and a constant rotational motion ω of the point P around C:

$$\left(\frac{dC\,(t)}{dt}\right)_{camera} = V \tag{3a}$$

and

$$\left(\frac{dr\,(t)}{dt}\right)_{camera} = \omega \times r. \tag{3b}$$

If the translational velocity V is not changing in time then the solution of Eq. (3a) is

$$C\,(t) = C_0 + V \cdot t. \tag{4}$$

C_0 is the initial position of the body-fixed coordinate system.

Let us now derive the solution for the rotational motion, Eq. (3b). For constant angular velocity the solution of Eq. (3b) can be found in several ways. [3] We rewrite Eq. (3b) in a slightly different form

$$\left(\frac{dr\,(t)}{dt}\right) = \Omega \cdot r \tag{5a}$$

---

[3] I thank Mark Snyder for a suggestion that led me to this simple solution.

where

$$\Omega = \begin{pmatrix} 0 & -\omega_z & +\omega_y \\ +\omega_z & 0 & -\omega_x \\ -\omega_y & +\omega_x & 0 \end{pmatrix} \tag{5b}$$

is the matrix specifying the rotation of the body. The solution of Eq. (5a) is

$$r(t) = e^{\Omega t} \cdot r_0. \tag{6}$$

r_0 is the initial position of the point P in the body-fixed coordinate system. Here, we make an important observation that the matrix Ω satisfies the relation

$$\Omega^3 = - \parallel \omega \parallel^2 \cdot \Omega \tag{7a}$$

where $\parallel \omega \parallel$ is the norm of the angular velocity $\omega = (\omega_x, \omega_y, \omega_z)$

$$\parallel \omega \parallel^2 = \omega_x^2 + \omega_y^2 + \omega_z^2. \tag{7b}$$

This observation enables us to simplify the matrix equation (6). We find, using Eqs. (7), that the solution of Eq. (5) is

$$r(t) = \left[I + S(\omega, t) \cdot \Omega + C(\omega, t) \cdot \Omega^2 \right] \cdot r_0 \tag{8a}$$

where I is the identity matrix and we introduce shorthand notation

$$S(\omega, t) \equiv \frac{sin(\parallel \omega \parallel t)}{\parallel \omega \parallel}, \tag{8b}$$

$$C(\omega, t) \equiv \frac{(1 - cos(\parallel \omega \parallel t))}{\parallel \omega \parallel^2}. \tag{8c}$$

Eqs. (8) are similar to Rodrigues formula [19]. The use of functions $S(\omega, t)$ and $C(\omega, t)$ is particularly convenient for small rotations, when these functions are approximately independent of $\parallel \omega \parallel$.

The motion of a point P of a rigid body moving with constant translational velocity V and constant rotational velocity ω, with point's starting position R_0

$$R_0 = C_0 + r_0 \tag{9}$$

is then derived by combining the solutions for translational motion (Eq. (4)) and rotational motion (Eq. (8a)):

$$R(t) = R_0 + V \cdot t + \left[S(\omega, t) \cdot \Omega + C(\omega, t) \cdot \Omega^2\right] r_0. \tag{10}$$

Ω is given by Eq. (5b) and $S(\omega, t)$ and $C(\omega, t)$ are given by Eqs. (8b) and (8c).

Without loss of generality we can assume that $C_0 = 0$, i.e., that the center of rotation and the camera coordinate system coincide at time $t = 0$. [4] With the assumption $r_0 = R_0$ the equation Eq. (10) becomes:

$$R(t) = R_0 + V \cdot t + \left[S(\omega, t) \cdot \Omega + C(\omega, t) \cdot \Omega^2\right] R_0. \tag{11}$$

This equation describes helix-like trajectory of the point P moving with constant translational and rotational velocity.

There are 9 parameters in Eq. (11) three for each of: the initial position of the point R_0, the translational velocity V, and the rotational velocity ω. The non-linearity in ω is evident from the rotational (third) summand.

Let $Q = (x(t), y(t))$ be the central projection of the point P on the image plane. The components of $R(t) = (X(t), Y(t), Z(t))$ satisfy the following set of equations

$$fX(t) - x(t)Z(t) = 0 \tag{12a}$$

$$fY(t) - y(t)Z(t) = 0 \tag{12b}$$

where f is the focal length of the camera, assumed to be the unit of length ($f = 1$). We assume that $Z(t) \neq 0$, for all t. Since there are 9 unknown parameters we need at least 5 images to determine

[4] The search for parameters is slightly more difficult when this assumption is not used.

the motion parameters of a single point, P. (Each image supplies two equations for the unknown parameters.) Input parameters are image events $Q_i = (x_i, y_i, t_i) = (x(t_i), y(t_i))$ for some arbitrary times t_i, $i = 0, 1, \ldots, 4$. Some other combination of the number of images and the number of points belonging to the same rigid body can be used as long as there are enough equations to solve for the unknown parameters. However, a larger number of images gives a more reliable prediction of motion, and is thus preferred to a large number of points, unless there is a danger of feature disappearance during the time interval considered.

3 Method of Solution

In this section we use a generalized version of Newton's iterative method to solve the equations for the motion parameters which we developed in the previous section . We can foresee some of the advantages of the proposed solution, Eq. (11). The known type of non-linearity makes the method converge fast and be more stable. It is quite possible that there is an especially suitable numerical procedure for this type of non-linearity, although we have not found one yet.

The initial position of the point P is determined by the 3 parameters X_0, Y_0, and Z_0. Since Eqs. (12) are homogeneous equations of the first order in $X(t), Y(t)$, and $Z(t)$, the solution is determined up to a scale factor – usually referred to as the loss of depth during the central projection. We can set the scale factor equal to an arbitrary constant, Z_0, supplied in practice by some other method (e.g., from laser range data). The first image in the sequence then provides enough information to determine the two unknown parameters

$$X_0 = x_0 Z_0, \tag{13a}$$

$$Y_0 = y_0 Z_0. \tag{13b}$$

Thus, we are left with six unknown parameters labeled as a group

$$\xi \equiv (V_x, V_y, V_z;\ \omega_x, \omega_y, \omega_z) \tag{14}$$

for which we need at least three more images. In the work presented here we were not concerned with the determination of the structure of the moving body. One method for determining the structure can be found in the work by Broida and Chellappa [14]. Their method results in eight unknown parameters and computes the structure of the moving body.

The closed form of the solution enables us to simplify the formulation of the minimization problem in that the error becomes a linear function of point coordinates. Namely, assuming no over-determined system of equations, we can use a generalized Newton iterative procedure [37] in the form

$$J(\xi_n) = (\xi_{n+1} - \xi_n) \cdot \varepsilon(\xi_n), \qquad n = 0, 1, \ldots \tag{15}$$

where the error of the solution is

$$\varepsilon(\xi) = \begin{pmatrix} X(t_1) - x_1 Z(t_1) \\ X(t_2) - x_2 Z(t_2) \\ X(t_3) - x_3 Z(t_3) \\ Y(t_1) - y_1 Z(t_1) \\ Y(t_2) - y_2 Z(t_2) \\ Y(t_3) - y_3 Z(t_3) \end{pmatrix} \tag{16}$$

ξ_{n+1} is a new, better set of values for the unknown motion parameters. $J(\xi_n)$ is the Jacobian matrix

$$J_{ij}(\xi_n) = \frac{\partial \varepsilon_i}{\partial \xi_n^j}, \qquad i, j = 1, \ldots, 6 \tag{17}$$

which is easily computable from Eqs. (16). Note that since we already know the explicit dependence of the coordinates of P on the motion parameters, we do not have to minimize expressions of the form

$$\varepsilon_i' = \frac{f X(t_i)}{Z(t_i)} - x_i \tag{18}$$

often found in optic flow (differential) computations. By formulating the minimization problem in this way, Eqs. (16)), we avoid additional non-linearity of equations and are headed for faster and more stable convergence.

4 Experimental Results

In this section we give preliminary results for experimental runs designed to test the ideas and feasibility of our approach. The results demonstrate the generality of the method and its very fast convergence.

The algorithm is implemented in Common Lisp on a TI-Explorer. The algorithm takes only a few seconds to compute motion parameters. The sequence of images in Figure 2 shows centrally-projected trajectories after each step in Newton's iterative procedure for a point moving with constant linear and angular velocity. In each of the images in Figure 2 there are three different types of trajectories. One, labeled with squares, represents the correct (goal) trajectory. Each square is centered around the image coordinates (x_i, y_i), of the point moving with the correct set of motion parameters $\xi \equiv (V;\ \omega)$. Image events $Q_i = (x_i, y_i, t_i)$, $i = 0, 1, 2, 3$, are provided as input parameters to the iterative algorithm. An initial guess for motion parameters ξ_0 produces the trajectory labeled with crosses. The trajectory which represents the motion with improved motion parameters ξ_s, $s = 1, 2, 3, \ldots$ (after each step s of iteration) is labeled by triangles. Note that the initial position of the point R_0 is found, according to Eqs. (13), from the first pair of image coordinates (x_0, y_0). This is the reason that all the trajectories start from the same initial point (labeled as "0").

To demonstrate the generality of the method, we choose to detect motion of an arbitrary linear and angular velocity. Figure 2 illustrates experiments which detect motion of a point on a body moving with linear velocity $V = (0.1, 0.2, 0.3)$ and angular velocity $\omega = (0.3, -0.2, 0.2)$, i.e.,

$$\xi_\infty = (0.1, 0.2, 0.3;\ 0.3, -0.2, 0.2).$$

The unit of length is $f = 1$ and the unit of time is 1.

Convergence to the correct solution is highly dependent on the choice of the initial guess. Even a very rough and easily obtainable estimate of the motion parameters (e.g., the sign of the z-component of the linear and angular velocity) drastically reduces the amount of search. As can be seen in Figure 2, our initial guess $\xi_0 = (0.0, 0.3, 0.4;\ 0.2, 0.0, 0.2)$, is "far" from the correct solution

$$\|\xi_0 - \xi_\infty\| / \|\xi_\infty\| = 50.8\%,$$

but has the same sign of the curvature and direction of motion in one direction. After the first iteration we get

$$\textit{step } 1: \quad \xi_1 = (0.21, 0.42, 0.48;\ 0.50, -0.36, 0.25). \qquad (\textit{Figure 2a})$$

This is shown as the trajectory labeled with triangles in Figure 2a. In each of the subsequent figures we show the current trajectory (triangles) after the second, third, and fourth iteration, together with the initially guessed, and the correct, trajectory.

$$\textit{step } 2: \quad \xi_2 = (0.07, 0.25, 0.47;\ 0.33, -0.25, 0.25), \qquad (\textit{Figure 2b})$$

$$\textit{step } 3: \quad \xi_3 = (0.09, 0.21, 0.30;\ 0.31, -0.20, 0.21), \qquad (\textit{Figure 2c})$$

$$\textit{step } 4: \quad \xi_4 = (0.099, 0.201, 0.300;\ 0.300, -0.200, 0.201). \qquad (\textit{Figure 2d})$$

Thus, after only 4 iterations the relative error in the motion parameters is

$$\|\xi_4 - \xi_\infty\| / \|\xi_\infty\| = 0.3\%.$$

Note, in particular, that components V_x and ω_y were first initialized to 0 but are still correctly computed. Good convergence was also found for many other types of motions and initial guesses.

5 Discussion

As expected for any set of non-linear equations, there are initial guesses for which the convergence is poor and the expected solution is not found. The situation can be compared with finding the roots of the equation $z^3 - 1 = 0$ in the complex plane. Depending on the initial guess any of the three roots can be found. Although the solution is usually the root which is the closest to the initial guess, there are some cases where a different root is found.

The problem can be sometimes resolved using over-constrained systems (more image events than necessary). We are currently working on exploring the relation between the type of motion and the value of the initial guess, in order to devise appropriate initial guessing schemes.

In a forthcoming paper we will present a detailed analysis of the algorithm for both synthetic and real images, and for several types of motion. We are presently exploring a number of important issues, before applying the method to a sequence of real images. One problem is the feature correspondence over several images. A larger distance between image features will better define the trajectory and recover motion parameters more accurately. Unfortunately, distant features are more difficult to correlate between images. A strength of our method is that it allows arbitrary time intervals between successive images. Allowing larger time intervals between images facilitates the separation of different motions with similar projections.

Another issue is the sensitivity of the solution to the positional error of the image features. The idea of smoothing [30] applied to the time coordinate, together with error analysis similar to [14,26,35] could be used as a promising start. The Newton iterative method is easily generalized to over-determined systems, so as to insure greater robustness to noise. A related problem is the non-uniqueness of solutions of the non-linear equations. This is the case, for example, when the trajectory has a much larger $\| \omega \|$ than expected, but passes through the same space-time events, a phenomenon that could be called the "stroboscopic" effect. [5]. In a sense it is an undersampling in the time domain, that could be corrected. Over-determined systems have a potential to solve this problem as well. If we can assume that angular velocity is small, then the solution space of the iteration procedure can be restricted and multiple solutions can be avoided.

[5]remark by Yigal Gur

If the rotational frequencies are very high, a completely different problem arises: it becomes increasingly hard to track the motion of a single feature due to its periodic disappearance. Our approach is suitable for high frequencies since it does not require the constant time interval between the time frames. However, some estimation techniques of the feature positions will have to be incorporated.

We also plan to investigate the motion of several objects simultaneously, as well as that of several features on a single object. Tracking of several features on a single object can produce an estimate of the 3-D structure of the object or it can be used to facilitate motion segmentation. Another very interesting direction for investigation is the derivation of equations (in a manner similar to the one presented in this paper) for structures more complex than a point, i.e., to lines, planes, and other surfaces of a rigid body [25,36,40]. Finally, using perturbation techniques these equations can be generalized to motions of bodies with slowly changing shape, and/or slowly changing motion parameters.

5.1 Conclusion

In conclusion we have used a spatio-temporal analysis of image events to present a quite general, robust and computationally efficient method for the recovery of motion parameters of moving objects under the formulation posed by Shariat [34] of constant motion. The closed form solution (containing exact and explicit information about the body motion) has other advantages: solution strategies can be adapted to the known type of non-linearity resulting in faster and more reliable convergence; the initial guessing scheme can exploit constraints derived from the proposed solution; a generalization to more images, more features, and/or variable time intervals between images is readily available; and, finally, the constraint of constant motion parameters can be relaxed, using perturbation techniques for slowly varying parameters.

ACKNOWLEDGMENTS

I would like to thank Edward Riseman for encouragement and careful readings of this article. Thanks also go to Jasmina Pavlin for careful readings of this paper and useful suggestions, and to Allen Hanson, Mark Snyder, P. Anandan and Deborah Strahman for useful comments.

REFERENCES

[1] G. Adiv, *Determining 3-D Motion and Structure from Optical Flow Generated by Several Objects.* Proceedings of Image Understanding Workshop, DARPA, October 1984, pp. 113-129.

[2] J. K. Aggarwal, and A. Mitiche, *Structure and Motion from Images.* Proceedings of Image Understanding Workshop, DARPA, Miami Beach, Florida, December 1985, pp. 89-97.

[3] P. Anandan, *A Computational Framework and an Algorithm for the Measurement of Visual Motion.* To appear in the International Journal on Computer Vision, 1988.

[4] P. Anandan, *Measuring Visual Motion From Image Sequences.* Ph.D. Dissertation and Technical Report 87-21, Department of Computer and Information Science, University of Massachusetts, Amherst, March 1987.

[5] P. Anandan, *A Unified Perspective on Computational Techniques for the Measurement of Visual Motion.* Proceedings of the International Conference on Computer Vision, London, England, June 1987.

[6] P. Anandan, *Computing Optical Flow From Two Frames of an Image Sequence.* Technical Report 86-16, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts 01003, April 1986.

[7] D. H. Ballard, and O. A. Kimball, *Rigid Body Motion from Depth and Optical Flow.* Computer Vision, Graphics, and Image Processing, **22**, 1983, pp. 95-115.

[8] S. Bharwani, E. Riseman, and A. Hanson, *Refinement of Environment Depth Maps Over Multiple Frames.* Proceedings of Image Understanding Workshop, DARPA, December 1985 pp. 413-420.

[9] S. Bharwani, E. Riseman, and A. Hanson, *Refinement of Environmental Depth Maps Over Multiple Frames.* Proceeding of the IEEE Workshop on Motion: Representation and Analysis, Charleston, SC, May 1986.

[10] R. C. Bolles and H. H. Baker, *Epipolar-Plane Image Analysis: A Technique for Analyzing Motion Sequences.* IEEE Proceedings of the 3rd Workshop on Computer Vision: Representation and Control, Bellaire, Michigan, October 1985, pp. 168-178.

[11] M. Boldt and R. Weiss, *Token-Based Extraction of Straight Lines.* To appear in the International Journal on Computer Vision, 1988.

[12] S. M. Bozic, *Digital and Kalman Filtering.* Edward Arnold, London, 1979.

[13] T. J. Broida and R. Chellappa, *Estimation of Object Motion Parameters from Noisy Images.* Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisko, California, June 1985, pp. 90-99.

[14] T. J. Broida and R. Chellappa, *Kinematics of a Rigid Object from a Sequence of Noisy Images: A Batch Approach.* Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1986, pp. 176-182.

[15] A. L. Fetter and J. D. Walecka, *Theoretical Mechanics of Particles and Continua.* McGraw-Hill, Inc., New York, 1980.

[16] J. J. Gibson, *The Perception of the Visual World.* Riverside, Cambridge, Mass, 1950.

[17] E. C. Hildreth, *Computation of the Velocity Field.* Proc. R. Soc. Lond., 1984, **B 221**, pp. 189-220.

[18] B. K. P. Horn and B. G. Schunck, *Determinining Optical Flow.* MIT, A.I. Memo 572, April 1980. Also, Artificial Intelligence, 1981, **17**, pp. 185-203.

[19] B. K. P. Horn, *Robot Vision.* The MIT Press, Cambridge, Massachusetts, 1986, p. 450.

[20] A. H. Jazwinski, *Stochastic Processes and Filtering Theory.* Academic Press, New York, 1970.

[21] G. Johansson, *Spatio-Temporal Differentiation and Integration in Visual Motion Perception.* Psych. Res., **38**, 1976, pp. 379-383.

[22] J. J. Koenderink and A. J. van Doorn, *Local Structure of Movement Parallax of the Plane.* J. Opt. Soc. Am. **66**, No. 7, July 1976, pp. 717-723.

[23] D. T. Lawton, *Processing Dynamic Image Sequences from a Moving Sensor.* Ph. D. Dissertation. University of Massachusetts, Amherst, MA 01003. Also, Technical Report 84-05, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts 01003, February 1984.

[24] H. C. Longuet-Higgins and K. Prazdny, *The Interpretation of a Moving Retinal Image.* Proc. R. Soc. Lond., 1980, **B 208**, pp. 385-397.

[25] D. G. Lowe, *Three-Dimensional Object Recognition form Single Two-Dimensional Images.* Technical Report No. 202, Courant Institute, New York, February 1986.

[26] D. H. Marimont, *Inferring Spatial Structure from Feature Correspondences.* Ph.D. Dissertation, Stanford University, March 1986.

[27] P. S. Maybeck, *Stochastic Models, Estimation and Control.* Vols. 1-3. Academic Press, New York, 1982.

[28] G. Medioni and R. Nevatia, *Matching Images Using Linear Features.* IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-6**, November 1984, pp. 675-685.

[29] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics, II.* McGraw-Hill, New York, 1953.

[30] I. Pavlin, E. Riseman, and A. Hanson, *Translational Motion Algorithm with Global Feature Constraints.* Technical Report 86-58, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts 01003, December 1986.

[31] K. Price, *Relaxation Matching Techniques - A Comparison.* IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-7**, 1985, pp. 617-623.

[32] J. W. Roach and J. K. Aggarval, *Determining the Movement of Objects from a Sequence of Images.* IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-2**, November 1980, pp. 554-562.

[33] I. K. Sethi, and R. Jain, *Finding Trajectories of Feature Points in a Monocular Image Sequence.* **PAMI-9**, January 1987, pp. 56-73.

[34] H. Shariat, *The Motion Problem: How to use more than Two Frames.* Ph. D. Dissertation. University of Southern California, Los Angeles. Also, Technical Report IRIS 202, Institute for Robotics and Intelligent Systems, University of Southern California, Los Angeles, CA 90089-0273, October 1986.

[35] M. Snyder, *The Accuracy of 3D Parameters in Correspondence-Based Techniques.* Technical Report 86-28, Department of Computer and Information Science, University of Massachusetts, Amherst, Ma 01003, July 1986.

[36] M. E. Spetsakis and J. Aloimonos, *Closed Form Solution to the Structure From Motion Problem from Line Correspondences*, Proc. AAAI-87, Seatle, WA, July 1987, pp. 738-743.

[37] F. Stummel and K. Hainer, *Introduction to Numerical Analysis.* Scottish Academic Press, Edinburgh, 1980.

[38] R. Y. Tsai and T. S. Huang, *Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch.* Proc. PRIP, 1981, pp. 94-97.

[39] A. M. Waxman and S. Ullman, *Surface Structure and 3-D Motion from Image Flow: A Kinematic Analysis.* Technical report CS-TR-1332, Carnegie-Mellon, October 1983.

[40] L. Williams and A. Hanson, *Translating Optical Flow Into Token Matches.* To appear in DARPA Image Understanding Workshop, Cambridge, MA, April 1988.

[41] H. J. Wünshe, *Detection and Control of Mobile Robot Motion by Real-Time Computer Vision.* SPIE Conference 727 on "Mobile Robots", Cambridge, MA, October 1986.

[42] B. L. Yen and T. S. Huang, *Determining 3-D Motion and Structure of a Rigid Body Using the Spherical Projection.* Computer Vision, Graphics, and Image Processing, **21**, 1983, pp. 21-32.

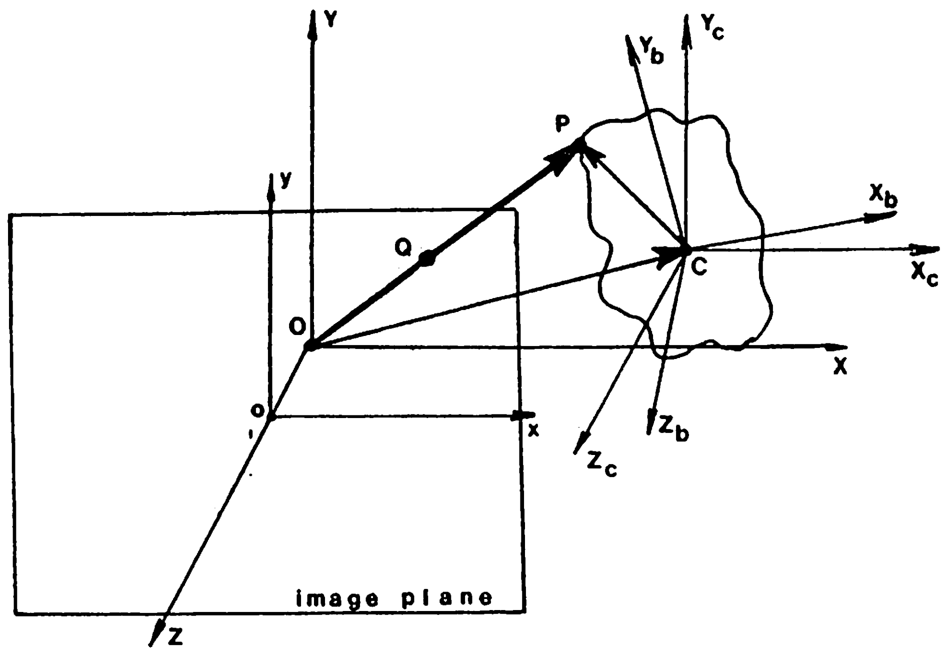
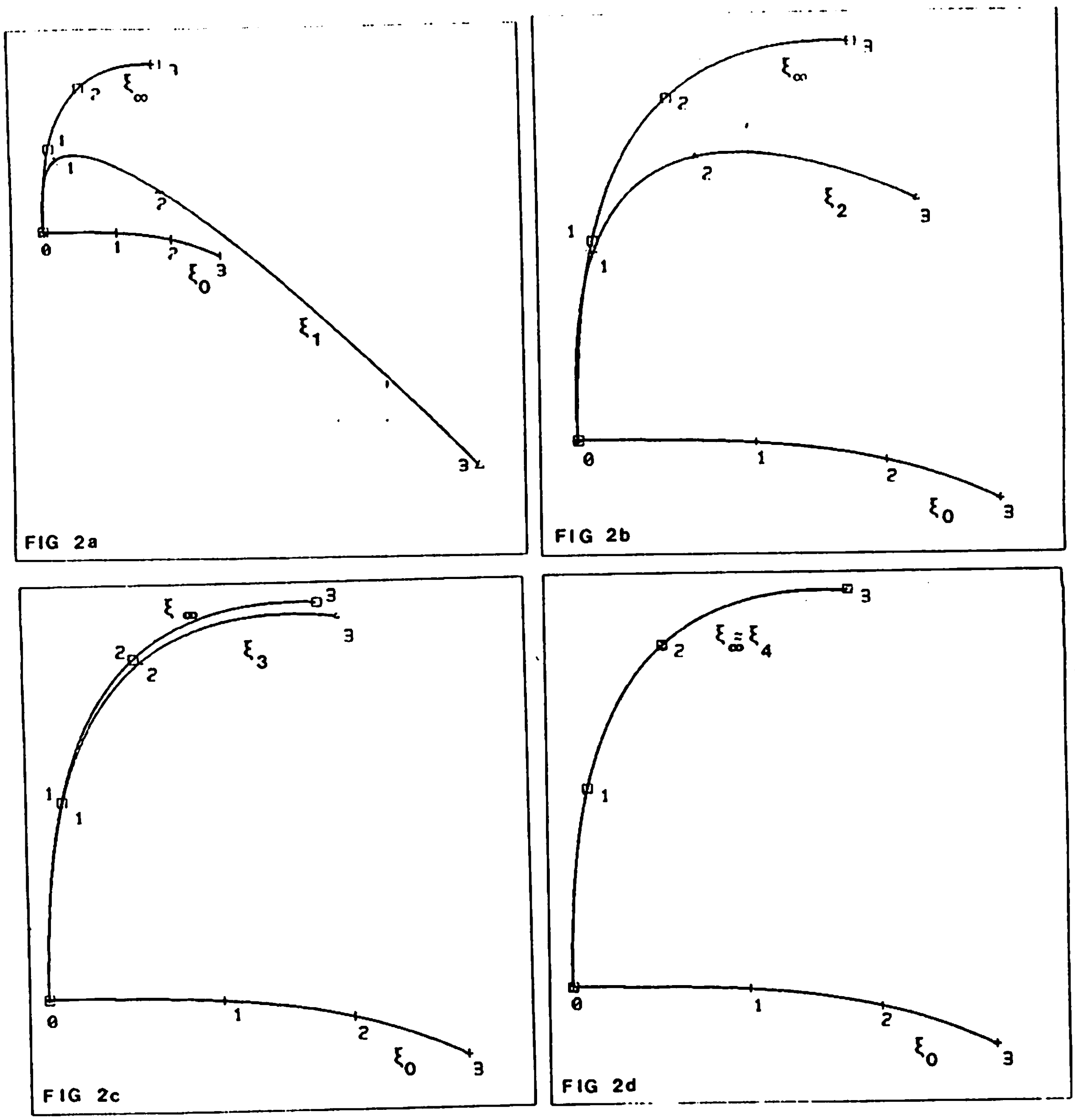


Figure 1: Motion of arbitrary rigid body, and coordinate systems.



Four iterative steps in the recovery of motion parameters for the trajectory labeled by squares. Crosses label initially guessed trajectory ξ_0, triangles label trajectories after the step s (trajectories ξ_s, $s = 1,2,3,4$) of the Newton iteration method described in text . Squares label the correct solution ξ_∞.

Figure 2: Convergence of Iterative Method in Recovery of Motion Parameters.