Estimation Of General Rigid Body Motion From A Long Sequence Of Images

MS-CIS-90-22
GRASP LAB 204

Siu-Leong Iu
Kwangyoen Wohn

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104-6389

April 1990

ESTIMATION OF GENERAL RIGID BODY MOTION FROM A LONG SEQUENCE OF IMAGES

Siu-Leong Iu & Kwangyoen Wohn

The Grasp Laboratory
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104.

Abstract

In estimating the 3-D rigid body motion and structure from time-varying images, most of previous approaches which exploit a large number of frames assume that the rotation, and the translation in some case, are constant. For a long sequence of images, this assumption in general is not valid. In this paper, we propose a new state estimation formulation for the general motion in which the 3-D translation and rotation are modeled as the polynomials of arbitrary order. Extended Kalman filter is used to find the estimates recursively from noisy images. A number of simulations including the Monte Carlo analysis are conducted to illustrate the performance of the proposed formulation.

Key words: Image motion analysis, structure from motion, extended Kalman filter.

Acknowledgement: This research was supported in part by DARPA grant NOOO14-88-K-0632, NSF equipment grant CCR-8716975 and NSF grant IRI89-06770.

1. INTRODUCTION

Token-matching approach and optical flow approach are two kinds of approaches to the recovery of 3-D motion and structure from images by using, respectively, the measurement of projective position [Agga81, Lin86, Seth87, Fang84, Rana80] and the measurement of optical flow [Ullm81, Horn81, Wohn83, Hild83, Nage83, Schu85, Heeg86]. Our work described in this paper belongs to the token-matching approach. For the token matching approach, point features in the scene have been studied extensively [Ullm79, Roac80, Nage81, Long81, Huan81, Tsai84, Faug87, Nage86]. Other features like line segments and conic arc in the scene have also been used [Liu86, Miti86, Faug87, Tsai83]. For the optical flow approach, 3-D motion is determined from the measurement of optical flows, and its temporal and spatial derivatives [Long80, Waxm85, Long84, Subb85, Kana85, Waxm86, Wu86]. Aggarwal and Nandhakumar [Agga88] gave an excellent and up to date review of the whole field of estimating the 3-D structure and motion from sequences of monocular and stereoscopic images.

Most of existing *structure-from-motion* algorithms which utilize a small number (two or three) of frames perform poorly under the presence of noise in the measurement. Recently, a number of researchers [Weng87, Broi86a,b, Boll85, Iu89a, Broi89, Kuma89] have proposed to use a large number of frames in order to improve the estimation performance. The very first step towards the "multi-frame" analysis is to model the kinematic variables of moving object during an extended period of time. As far as the translational motion is concerned, the trajectory w.r.t. time describes the motion completely. Since an object can move almost arbitrarily in front of a camera, we do not know this trajectory explicitly. However, according to Weierstrass approximation theorem, we can approximate any continuous trajectory over a closed and bounded time interval by using polynomials. In practice, we can only use low order polynomials because of the numerical stability and the real-time constraints. If the object moves in the 3-D space smoothly, such low order polynomials may be good enough in describing its motion locally. Note that the projectile motion of a throwing ball can be represented by using second order polynomials. In the following we contrast our approach to the two previous work which is considered as the major contribution to our understanding of motion estimation from a long sequence of images.

Weng, Huang and Ahuja [Weng87] proposed the locally constant angular momentum model in estimating 3-D motion from the measurements of projective positions. Their model assumed that angular momentum was constant over a short time interval, the moving body possessed an axis of symmetry and the motion of the rotation center was approximated as a polynomial. The first two assumptions are required in their derivation because they wanted the Euler's equations integrable. In our

approach, we do not make these two assumptions because we directly derive the relation between the unknown motion parameters and the projective positions, rather than solving the equations involving the external torque and the angular momentum. In finding the 3-D motion, they first estimated the rotation matrices and the translation between the frames by using "two-frame" motion analysis, and then determined the motion of the rotation center, i.e., their approach considers each frame separately. In our approach, we utilize all the available information in the temporal and spatial domains to estimate the translation velocity of the rotation center, the rotation of the rigid body and the relative depth simultaneously. Furthermore, in their paper, only simulations on binocular image sequences were reported.

Broida and Chellappa [Broi86a,b] estimated motion parameters sequentially from the projective position of multiple points in a sequence of noisy images. They used a dynamic model to describe the temporal behavior of parameters and employed the iterative extended Kalman filter to estimate them. In [Broi86a], they estimated the motion of a two-dimensional object from its one-dimensional projections. They assumed that the image coordinates of feature points were available, that the motion was unforced, that the absolute distance to the center of rotation was known and that the noise level in the Kalman filter formulation was known. In [Broi86b], they extended their work to 3-D rigid body motion. Quaternions were used to describe the rotation. Motion was modeled by a truncated Taylor series but only the linear term was used in their derivation. Simulations on very special motion only, such as pure translation or rotation about a fixed and known axis, were reported in their paper. In [Broi86c], they used batch approach to find the estimate. Their result on uniqueness of rigid body motion was limited to the pure translational motion [Broi89]. For the non-constant rotation, numerical integration was proposed [Broi86b,c, Broi89] but neither specific algorithms nor experiments were given.

Iu and Wohn [Iu89a] argued that the estimation of motion should exploit the temporal information from images first rather than seeking for the motion constraint from the spatial domain. They showed that the 3-D velocity of a single point up to a scalar factor could be recovered from images and proved the uniqueness of solution for the case that the 3-D velocity is modeled as an arbitrary order of polynomials. Regression relation between unknown motion parameters and measurements from noisy images was derived and the batch approach was used to find the optimal estimate under the criterion of Maximum Likelihood. They also extended their work to 3-D rigid body motion with constant angular velocity. In [Iu89b], extended Kalman filter was used to estimate the motion sequentially.

We have observed that most of the previous approaches modeled the rotation, and the translation for some cases, as a constant. One may attempt to extend the analysis for the constant motion to the one for the general motion by expanding the translational and the rotational parameters in polynomials and plugging them into the original formulation of the constant motion. However, for the rotational motion, although we can still use a higher order polynomials to describe the rotation, the state equation relating the rotation and the 3-D trajectory of the feature points respect to the rotational center is time varying. Since there is no explicit closed form solution for this equation, the previous approaches were not able to handle the non-constant rotational motion effectively. In this paper, we propose a new state estimation formulation in which the explicit solution of the above state equation is not required. Consequently, the analysis can include the object motion with arbitrary orders of translation and rotation.

The rest of the paper is organized as follows. Section 2 describes the model of general rigid body motion. Section 3 gives the state estimation formulation. The motion with constant acceleration is discussed first. Section 4 gives the result of simulations. Some concern on generating the test data is also pointed out. Section 5 is the conclusion of this paper with discussion.

2. MODEL OF RIGID BODY MOTION

Suppose that there are n_p feature points on the visible surface of the rigid object and that we have measured the projected position of these points as the object moves in front of a camera. It is well known that any rigid body motion can be represented by the translation of the rotational center and the rotation of the entire object with respect to this rotational center. We also know that we can only recover the translation up to a scale factor because of the perspective projection. Thus, our objective is to find this translation of the rotational center up to the scalar factor and the rotation from the observed projected positions.

Let $\underline{P}_i(t) = [X_i(t)\ Y_i(t)\ Z_i(t)]^T$, $\underline{p}_i(t) = [x_i(t)\ y_i(t)]^T$ and $\underline{V}_i(t) = [V_{Xi}(t)\ V_{Yi}(t)\ V_{Zi}(t)]^T$ be the 3-D position, the projected position and the instantaneous velocity of i-th feature point at time t, respectively. The symbol 'T' denotes the transpose of a vector. Then we have

$$\frac{d}{dt}\underline{P}_i(t) = \underline{V}_i(t)\ , \tag{2.1}$$

$$x_i(t) = X_i(t) / Z_i(t), \quad y_i(t) = Y_i(t) / Z_i(t)\ . \tag{2.2a,b}$$

Note that we have used the pin-hole camera model and have scaled the Z-axis such that the focal length is normalized to one [Long80]. Let $\underline{\Omega}(t) = [\Omega_X(t)\ \Omega_Y(t)\ \Omega_Z(t)]^T$ be the angular velocity of the object respect to the rotational center $\underline{P}_0(t)$. From the assumption of rigidity, we have

$$\underline{V}_i(t) - \underline{V}_0(t) = \underline{\Omega}(t) \times [\ \underline{P}_i(t) - \underline{P}_0(t)\ ]\ , \tag{2.3}$$

where the symbol '×' denotes the cross product of vectors. We can express (2.3) alternatively in the following form:

$$\frac{d}{dt}\ [\ \underline{P}_i(t) - \underline{P}_0(t)\ ] = A(t)\ [\ \underline{P}_i(t) - \underline{P}_0(t)\ ]\ , \tag{2.4}$$

where $$A(t) = \begin{bmatrix} 0 & -\Omega_Z(t) & \Omega_Y(t) \\ \Omega_Z(t) & 0 & -\Omega_X(t) \\ -\Omega_Y(t) & \Omega_X(t) & 0 \end{bmatrix}\ . \tag{2.5}$$

If the object moves smoothly in the 3-D space, we may model the translation and the rotation as the following polynomials with orders d_t and d_r respectively.

$$\underline{P}_0(t) = \sum_{n=0}^{d_t} \underline{P}_0^{[n]}(t_0)\ \frac{(t - t_0)^n}{n!}\ , \tag{2.6}$$

$$\underline{\Omega}(t) = \sum_{n=0}^{d_r} \underline{\Omega}^{[n]}(t_0)\ \frac{(t - t_0)^n}{n!}\ , \tag{2.7}$$

where $\underline{P}_0^{[n]}(t_0) = [\ X^{[n]}(t_0)\ \ Y^{[n]}(t_0)\ \ Z^{[n]}(t_0)\ ]^T$ and $\underline{\Omega}^{[n]}(t_0) = [\ \Omega_X^{[n]}(t_0)\ \ \Omega_Y^{[n]}(t_0)\ \ \Omega_Z^{[n]}(t_0)\ ]^T$ are the n-th derivatives of $\underline{P}_0(t)$ and $\underline{\Omega}(t)$ evaluated at time $t = t_0$. Note that $\underline{P}_0^{[0]}(t_0) = \underline{P}_0(t_0)$ and $\underline{\Omega}^{[0]}(t_0) = \underline{\Omega}(t_0)$. The physical interpretation of $\underline{P}_0^{[1]}(t_0)$ and $\underline{P}_0^{[2]}(t_0)$ are the velocity and acceleration of translation at time t_0, respectively. The $\underline{\Omega}^{[0]}(t_0)$ and $\underline{\Omega}^{[1]}(t_0)$ are the angular velocity and acceleration of rotation at time t_0, respectively. Then our goal is to estimate the translational coefficients $\underline{P}_0^{[n]}(t_0)$, $n = 0, 1, \cdots, d_t$ in (2.6) scaled by the factor representing the absolute depth and rotational coefficients $\underline{\Omega}^{[n]}(t_0)$, $n = 0, 1, \cdots, d_r$ in (2.7), from the measurement of $\underline{p}_i(t_j)$, $i = 0, \cdots, n_p - 1$, $j = 0, 1, \cdots$.

If the rotation is constant, i.e., $\underline{\Omega}(t) = \underline{\Omega}(t_0)$, then A(t) is a constant matrix and the state equation (2.4) has the close-form solution

$$\underline{P}_i(t) - \underline{P}_0(t) = e^{A\,[\,t - t_0\,]}\ [\ \underline{P}_i(t_0) - \underline{P}_0(t_0)\ ]\ , \tag{2.8}$$

where $e^{A\,[\,t - t_0\,]}$ is the state transition matrix and e^{At} is equal to the inverse Laplace transform of $(sI - A)^{-1}$. Then the rigid body motion can be determined by solving the regression problem [Iu89a] or the state estimation problem [Broi86b]. Unfortunately, if the rotation is time varying, there is in general no closed-form expression for $[\ \underline{P}_i(t) - \underline{P}_0(t)\ ]$ in terms of $\underline{\Omega}(t)$ and $[\ \underline{P}_i(t_0) - \underline{P}_0(t_0)\ ]$ because of the time-varying nature of A(t) [Iu89a]. This is why one can not extend the derivation of the previous approaches [Iu89a, Broi86b] to the general motion in a straightforward manner. In this paper, we develop a state estimation formulation in which the explicit solution of (2.4) is not required, given an arbitrary order of rigid body motion.

3. STATE ESTIMATION FORMULATION AND RECURSIVE SOLUTION

The problem of recovering general rigid body motion will be formulated as the state estimation problem. Let $\underline{s}(t)$ be the state vector which is composed of the unknown motion parameters. Let $\underline{m}(t_j)$ be the measurement vector which is formed by the available measurements from the image at time t_j;

$$\underline{m}(t_j) = [\ x_0(t_j)\ \ y_0(t_j)\ \ x_1(t_j)\ \ y_1(t_j)\ \cdots\ x_{n_p-1}(t_j)\ \ y_{n_p-1}(t_j)\ ]^T + \underline{n}(t_j)\ , \tag{3.1}$$

where $\underline{n}(t_j)$ is a discrete-time random process describing the noise in the measurements. The components of $\underline{n}(t_j)$ are assumed to be independent and $\{\ \underline{n}(t_j), j = 0, 1, \cdots\ \}$ are assumed to be white zero mean Gaussian random vectors with common variance matrix. If we can derive the state evolution and the relation between the measurement and state vectors in the following plant equation and measurement equation,

$$\frac{d}{dt}\ \underline{s}(t) = \underline{f}(\underline{s}(t))\ , \tag{3.2}$$

$$\underline{m}(t_j) = \underline{h}(\underline{s}(t_j)) + \underline{n}(t_j) \ , \tag{3.3}$$

where $\underline{f}(.)$ and $\underline{h}(.)$ are nonlinear vector functions, then the solution of our rigid body motion problem can be obtained by solving the nonlinear state estimation problem. Extended Kalman filter, iterative extended Kalman filter and nonlinear filter are commonly used for solving this type of problem recursively with different computational complexity [Mayb82]. In the following, we will first consider a special case-the motion of constant acceleration with three non-collinear feature points-and then will extend the derivation to the general case.

3.1 Special Non-constant Motion

In this section, we consider a non-constant motion in which the accelerations of translation and rotation are constant, and there are only three non-colinear points on the rigid body, i.e., $d_t = 2$, $d_r = 1$ and $n_p = 3$. This implies that $\underline{P}_0^{[m]}(t) = \underline{\Omega}^{[n]}(t) = [0\ 0\ 0]^T$ for $m \geq 3$ or $n \geq 2$.

Let the state vector

$$\underline{s}(t) = \left[\frac{X_0}{Z_0} \ \frac{Y_0}{Z_0} \ \frac{X_0^{[1]}}{Z_0} \ \frac{Y_0^{[1]}}{Z_0} \ \frac{Z_0^{[1]}}{Z_0} \ \frac{X_0^{[2]}}{Z_0} \ \frac{Y_0^{[2]}}{Z_0} \ \frac{Z_0^{[2]}}{Z_0} \ \Omega_X \ \Omega_Y \ \Omega_Z \ \Omega_X^{[1]} \ \Omega_Y^{[1]} \ \Omega_Z^{[1]} \right.$$
$$\left. \frac{X_1}{Z_0} \ \frac{Y_1}{Z_0} \ \frac{Z_1}{Z_0} \ \frac{X_2}{Z_0} \ \frac{Y_2}{Z_0} \ \frac{Z_2}{Z_0} \right]_t^T , \tag{3.4}$$

where the subscript 't' denotes the vector evaluated at time t. Let s_i be the i-th component of $\underline{s}(t)$ at time t. The total number of states, n_s is 20. The evolution of the first 14 components of the state vector which are related to the object translation and rotation can be found as follows.

$$\frac{d}{dt}\begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \\ s_8 \end{bmatrix} = \begin{bmatrix} s_3 - s_1 s_5 \\ s_4 - s_2 s_5 \\ s_6 - s_3 s_5 \\ s_7 - s_4 s_5 \\ s_8 - s_5^2 \\ - s_6 s_5 \\ - s_7 s_5 \\ - s_8 s_5 \end{bmatrix}, \quad \frac{d}{dt}\begin{bmatrix} s_9 \\ s_{10} \\ s_{11} \\ s_{12} \\ s_{13} \\ s_{14} \end{bmatrix} = \begin{bmatrix} s_{12} \\ s_{13} \\ s_{14} \\ 0 \\ 0 \\ 0 \end{bmatrix} . \tag{3.5a,b}$$

Next, we consider the relative position of feature point i w.r.t. the center of rotation. From (2.4) and using (2.2), we can find the time derivative of $[X_i(t)\ Y_i(t)\ Z_i(t)\ ]^T$ scaled by $1/Z_0(t)$ as follows.

$$\frac{1}{Z_0(t)} \frac{d}{dt}\begin{bmatrix} X_i(t) \\ Y_i(t) \\ Z_i(t) \end{bmatrix} = \begin{bmatrix} X_0^{[1]}(t) / Z_0(t) \\ Y_0^{[1]}(t) / Z_0(t) \\ Z_0^{[1]}(t) / Z_0(t) \end{bmatrix} + A(t) \begin{bmatrix} X_i(t) / Z_0(t) - x_0(t) \\ Y_i(t) / Z_0(t) - y_0(t) \\ Z_i(t) / Z_0(t) - 1 \end{bmatrix} . \tag{3.6}$$

Consequently, the evolution of the state components related to point i is obtained as follow. Let $p = 15 + 3(i - 1)$ for $i = 1, 2$,

$$\frac{d}{dt}\begin{bmatrix} s_p \\ s_{p+1} \\ s_{p+2} \end{bmatrix} = \begin{bmatrix} s_3 \\ s_4 \\ s_5 \end{bmatrix} + \begin{bmatrix} 0 & -s_{11} & s_{10} \\ s_{11} & 0 & -s_9 \\ -s_{10} & s_9 & 0 \end{bmatrix} \begin{bmatrix} s_p - s_1 \\ s_{p+1} - s_2 \\ s_{p+2} - 1 \end{bmatrix} - \begin{bmatrix} s_p \\ s_{p+1} \\ s_{p+2} \end{bmatrix} s_5 \,. \tag{3.7}$$

So far, we have obtained the plant equation which describes all the states changing in time. The measurement vector for the motion of interest is given by

$$\underline{m}(t) = [\, x_0(t) \;\; y_0(t) \;\; x_1(t) \;\; y_1(t) \;\; x_2(t) \;\; y_2(t) \,]^T + \underline{n}(t) \tag{3.8a}$$

$$= \underline{h}(\underline{s}(t)) + \underline{n}(t) \,. \tag{3.8b}$$

The total number of measurements at each instant is 6. From the definition of the state vector in (3.4) and using (2.2), the components of $\underline{h}(\underline{s}(t))$ at time t, h_i, are given by

$$h_1 = s_1, \quad h_2 = s_2 \,. \tag{3.9a,b}$$

Let $p = 15 + 3(i - 1)$ and $q = 1 + 2i$ for $i = 1, 2$, then

$$h_q = s_p / s_{p+2}, \quad h_{q+1} = s_{p+1} / s_{p+2} \,. \tag{3.10a,b}$$

Equations (3.9) and (3.10) form the measurement equation we sought for.

3.2 General Motion

Consider n_p points on the rigid body. The translation and the rotation are modeled as (2.6) and (2.7), respectively. Define the state vector as

$$\underline{s}(t) = \Bigg[\frac{X_0}{Z_0} \;\; \frac{Y_0}{Z_0} \;\; \frac{X_0^{[1]}}{Z_0} \;\; \frac{Y_0^{[1]}}{Z_0} \;\; \frac{Z_0^{[1]}}{Z_0} \;\; \cdots \;\; \frac{X_0^{[d_t]}}{Z_0} \;\; \frac{Y_0^{[d_t]}}{Z_0} \;\; \frac{Z_0^{[d_t]}}{Z_0} \;\; \Omega_X^{[0]} \;\; \Omega_Y^{[0]} \;\; \Omega_Z^{[0]} \;\; \cdots$$

$$\Omega_X^{[d_r]} \;\; \Omega_Y^{[d_r]} \;\; \Omega_Z^{[d_r]} \;\; \frac{X_1}{Z_0} \;\; \frac{Y_1}{Z_0} \;\; \frac{Z_1}{Z_0} \;\; \cdots \;\; \frac{X_{n_p-1}}{Z_0} \;\; \frac{Y_{n_p-1}}{Z_0} \;\; \frac{Z_{n_p-1}}{Z_0} \Bigg]_t^T \,. \tag{3.11}$$

The total number of states n_s is given as $3(d_t + d_r + n_p) + 2$. The components $\{s_k\}_{k=1}^{n_s}$ of the state vector at time t change with time as follows. For s_1 and s_2,

$$\frac{d}{dt}\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} s_3 \\ s_4 \end{bmatrix} - \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} s_5 \,. \tag{3.12}$$

For the states which represent the translational motion, let $p = 3 + 3(i - 1)$ for $i = 1, \cdots, d_t - 1$,

$$\frac{d}{dt}\begin{bmatrix} s_p \\ s_{p+1} \\ s_{p+2} \end{bmatrix} = \begin{bmatrix} s_{p+3} \\ s_{p+4} \\ s_{p+5} \end{bmatrix} - \begin{bmatrix} s_p \\ s_{p+1} \\ s_{p+2} \end{bmatrix} s_5 , \quad \frac{d}{dt}\begin{bmatrix} s_{3d_t} \\ s_{3d_t+1} \\ s_{3d_t+2} \end{bmatrix} = - \begin{bmatrix} s_{3d_t} \\ s_{3d_t+1} \\ s_{3d_t+2} \end{bmatrix} s_5 . \tag{3.13}$$

For the states which represent the rotational motion, let $p = (3d_t + 3) + 3i$ for $i = 0, \cdots, d_r - 1$,

$$\frac{d}{dt}\begin{bmatrix} s_p \\ s_{p+1} \\ s_{p+2} \end{bmatrix} = \begin{bmatrix} s_{p+3} \\ s_{p+4} \\ s_{p+5} \end{bmatrix} , \quad \frac{d}{dt}\begin{bmatrix} s_{3d_t+3d_r+3} \\ s_{3d_t+3d_r+4} \\ s_{3d_t+3d_r+5} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} . \tag{3.14}$$

For the states which represent the location of feature points, let $p = (3d_t + 3d_r + 6) + 3(i - 1)$ for $i = 1, \cdots, n_p - 1$,

$$\frac{d}{dt}\begin{bmatrix} s_p \\ s_{p+1} \\ s_{p+2} \end{bmatrix} = \begin{bmatrix} s_3 \\ s_4 \\ s_5 \end{bmatrix} + \begin{bmatrix} 0 & -s_{3d_t+5} & s_{3d_t+4} \\ s_{3d_t+5} & 0 & -s_{3d_t+3} \\ -s_{3d_t+4} & s_{3d_t+3} & 0 \end{bmatrix} \begin{bmatrix} s_p - s_1 \\ s_{p+1} - s_2 \\ s_{p+2} - 1 \end{bmatrix} - \begin{bmatrix} s_p \\ s_{p+1} \\ s_{p+2} \end{bmatrix} s_5 . \tag{3.15}$$

Using the definitions of state vector and measurement vector in (3.11) and (3.1), we can obtain the measurement equation in the form of (3.3) with the following components of $\underline{h}(\underline{s}(t_j))$, $\{h_k\}_{k=1}^{n_m}$, where n_m is the total number of measurements at each time and equals to $2n_p$.

$$h_1 = s_1, \quad h_2 = s_2 . \tag{3.16a,b}$$

Let $p = (3d_t + 3d_r + 6) + 3(i - 1)$ and $q = 1 + 2i$ for $i = 1, \cdots, n_p - 1$,

$$h_q = s_p / s_{p+2}, \quad h_{q+1} = s_{p+1} / s_{p+2} . \tag{3.17a,b}$$

In summary, we obtain the state equation which is formed by (3.12)-(3.15) and the measurement equation which is given by (3.16)-(3.17) for the general rigid body motion. Thus, the estimation of the general motion for a rigid body can be obtained by solving the nonlinear state estimation problem as we claimed earlier.

4. SIMULATION RESULTS

In order to illustrate the performance of the proposed approach to the motion estimation of the general rigid body motion, a number of experiments on simulated data are conducted. In the simulation, only three feature points are used. The focal length of the camera is set to one unit. The visible portion of

the image plane is (-0.36, 0.36) × (-0.36, 0.36) units. This portion corresponds to the viewing angle of ± 20 degrees. The observed image is considered as 256 × 256 pixels. We consider the motion of constant acceleration ($d_t = 2$ and $d_r = 1$). The time interval between frames is 0.04 second. To solve the nonlinear state estimation problem in (3.2) and (3.3) recursively, extended Kalman filter is used. The initial estimates of the translation and rotation are set to zero. The initial estimates of the related depth of the feature points are set to one.

On generating test data

The noisy measurements of the projected position $\underline{p}_i(t)$, $i = 0, \cdots, n_p-1$ at different sampling time t_j, $j = 0, 1, \cdots$, are generated by adding white zero mean Gaussian noise to the exact values of $\underline{p}_i(t)$. In the following, we explain and discuss the procedure of obtaining these exact projected positions from the given motion parameters: the 3-D position of feature points at time zero, the translational coefficients $\underline{P}_0^{[n]}(0)$, $n = 1, \cdots, d_t$ and rotational coefficients $\underline{\Omega}^{[n]}(0)$, $n = 1, \cdots, d_r$.

At a first glance, one may be tempted to use (3.12)-(3.15) to get the time evolution of state vector and then use (3.16) and (3.17) to obtain $\underline{p}_i(t)$ at different sampling times. However, since the time evolution of the states in (3.12)-(3.15) is nonlinear and coupled, a system of nonlinear differential vector equations with a large number of unknowns is required to solve. Although numerical techniques such as Runge Kutta method can be used to find this solution, the accumulated error may result the large error in the generated data. This is especially true when a large number of frames, say 100 frames, are involved. In order to overcome this problem, we use another approach. We first obtain the 3-D trajectory of rotational center $\underline{P}_0(t)$ at different sampling times by using (2.6) with known $\underline{P}_0^{[n]}(0)$, $n = 1, \cdots, d_t$. The 3-D trajectory of individual feature point, $\underline{P}_i(t)$, is determined by solving the state equation (2.4) and adding the result to $\underline{P}_0(t)$. Note that A(t) in (2.4) at any time can be found by using (2.7) with known $\underline{\Omega}^{[n]}(0)$, $n = 1, \cdots, d_t$. The state equation (2.4) is time-varying but it is not nonlinear. Also, there are only three unknowns in the equation. Runge Kutta method is used to solve the equation. Then the exact projected positions are obtained by using the perspective relation in (2.2).

Experiment 1: Various noisy levels

In this experiment, we compare the estimates at different noisy levels. The standard deviations of the noise are set to 0.5, 2.5 and 5 pixels. The 3-D translation $\underline{P}_0^{[1]}(t)$ is $[\,-1 \;\; 5 \;\; 10\,]^T + [\,-0.05 \;\; -2.5 \;\; -5\,]^T t$ units/second. The 3-D angular velocity $\underline{\Omega}(t)$ is $[\,-0.4 \;\; 0.5 \;\; 3\,]^T + [\,0.1 \;\; -0.3 \;\; -1.7\,]^T t$ radians/second. The 3-D positions of the three feature points at time zero are (4, -2, 20), (5, -5, 19.5) and (3, -5, 20.5)

units. Figure 1a shows the images of the simulated motion at every five frames. Figure 1b shows the exact and noisy trajectories of the first three feature points on the image plane (Only the trajectory with 2.5 pixel error is shown). Figures 1c and 1d are the x and y components of the trajectory, respectively. Figures 1e-1l are the exact and estimated $X_0^{[1]}(t)/Z_0(t)$, $Y_0^{[1]}(t)/Z_0(t)$, $Z_0^{[1]}(t)/Z_0(t)$, $\Omega_X(t)$, $\Omega_Y(t)$, $\Omega_Z(t)$, $Z_1(t)/Z_0(t)$ and $Z_2(t)/Z_0(t)$, respectively. From these results, we observed that the estimated errors decrease as the noise in the measurements decrease. The estimate errors for $X_0^{[1]} / Z_0(t)$ and $Y_0^{[1]} / Z_0(t)$ are smaller than that for $Z_0^{[1]} / Z_0(t)$. The estimate of rotations in X and Y direction are worst than that in Z direction. The estimate of relative depths can follow the change of the movement. As we use more frames, the estimate error decreases. For the case of 2.5 pixel error in the measurements, it takes about 60 frames to converge. This observation period is less than 2.5 seconds. Note that the estimate error is very large if we only use a smaller number of frames. This justifies that motion analysis using a small number of frames performs badly if there is noise in the measurements. We have similar observations discussed above for the following Monte Carlo analysis.

Experiment 2: Monte Carlo analysis

In this experiment, we run the simulation for fifty different sets of noise and compute the sample mean and standard deviation of the estimates . The standard deviation of the noise is 2.5 pixels. The 3-D translation $\underline{P}_0^{[1]}(t)$ is $[\,-3 \;\; 1 \;\; 10\,]^T + [\,1 \;\; -1 \;\; -2\,]^T\, t$ units/second. The 3-D angular velocity $\underline{\Omega}(t)$ is $[\,-0.5 \;\; 0.5 \;\; 2\,]^T + [\,0.2 \;\; -0.1 \;\; -1.4\,]^T\, t$ radians/second. The 3-D positions of the particles at time zero are (0, 0, 20), (4, -4, 19.5) and (-2, -4, 20.5) units. Figure 2a shows the images of the simulated motion at every five frames. Figures 2b, 2c and 2d show the exact and a typical noisy trajectories, and their x and y components. Figures 2e-2l show the sample mean and the $\pm$ 1 standard deviation of the estimated $X_0^{[1]}(t)/Z_0(t)$, $Y_0^{[1]}(t)/Z_0(t)$, $Z_0^{[1]}(t)/Z_0(t)$, $\Omega_X(t)$, $\Omega_Y(t)$, $\Omega_Z(t)$, $Z_1(t)/Z_0(t)$ and $Z_2(t)/Z_0(t)$ versus number of frames used in the estimation, respectively. It is observed that the sample mean converges to the true values and the sample standard deviations decrease as we use more frames.

5. CONCLUSION

We have proposed a new state formulation to analyze the object motion with arbitrary orders of translation and rotation from a sequence of video images. Extended Kalman filter was used to find the estimate recursively. The simulations showed that the proposed formulation was quite effective for estimating the non-constant rigid body motion, even the measurements had 5 pixels error. This formulation may be adopted with minor modification to the motion analysis for other measurements, such as the sequences of stereo images, spatially distributed infrared image, and radar observations of range

and bearings [Mayb82].

We may follow the derivations in this paper with modification to formulate the general rigid body motion in terms of quaternions [Broi86b]. However, we need to solve an additional time-varying 4×1 vector differential equation in order to obtain the quaternions at different instant from the non-constant angular velocity. Also, a non-linear algebra transformation is required to relate the estimated states to the measurements. These will increase the computational complexity and the additional error in the numerical integration.

For the model of rigid body motion we used, we assumed that the projective position of rotational center is visible, that the motion is smooth, and that the orders of the translation and rotation are known. Note that these assumptions are also used in [Weng87] and [Broi89]. For the first assumption, if the rotational center belongs to one of the feature points but we do not know which point is, one may regard each point as the rotational center, and apply the estimation procedure to each case. This approach is not unrealistic since the number of feature points we observed is commonly small and all the procedures can be processed simultaneously. Another approach is to detect the rotational center from the observed points and then estimate the motion. On the other hand, if the feature points do not contain the rotational center, then the method is not applicable directly.

In regard to the other two assumptions, since the object can move almost arbitrarily in front of a camera, the order of polynomials in describing the true motion may not match to the model exactly. Also, the object may change its motion abruptly. One such example is a ball bouncing off the ground. This situation leads us to consider three classes of model mismatch: undermodeling, overmodeling and parameter jumping. We have analyzed the performance degradation due to these model mismatches and proposed the *Finite Lifetime Alternately Triggered Multiple Model Filter* (FLAT MMF), as a new solution. This issue is dealt in [Iu89b, Iu90a].

Reference

[Agga81]
J.K. Aggarwal, L.S. Davis and W.N. Martin, "Correspondence process in dynamic scene analysis", Proc. of IEEE, 1981, pp. 562-572.

[Agga88]
J.K. Aggarwal and N. Nandhakumar, "On the computation of motion from sequence of image- a review", Proc. of IEEE, August 1988, pp. 917-935.

[Boll85]
R.C. Bolles and H.H. Bakers, "Epipolar-plane image analysis: a technique for analyzing motion sequences", in Workshop on computer vision: representation and control, Oct. 1985.

[Broi86a]
T.J. Broida and R.Chellappa, "Estimation of object motion parameters from noisy images", IEEE PAMI, January 1986 , pp. 90-99.

[Broi86b]
T.J. Broida and R. Chellappa, "Kinematics and structure of a rigid object from a sequence of noisy images", IEEE Workshop on visual motion 1986, pp. 95-100.

[Broi86c]
T.J. Broida and R. Chellappa, "Kinematics and structure of a rigid object from a sequence of noisy images: A batch approach", Proc. IEEE Conf. Computer Vision and Pattern Recognition, Miami Beach, FL, June 1986, pp. 176-182.

[Broi89]
T.J. Broida and R. Chellappa, "Experiments and uniqueness results on object structure and kinematics from a sequence of monocular images", IEEE Workshop on visual motion, March 1989, pp. 21-30.

[Fang84]
J.Q. Fang and T.S. Huang, "Solving three-dimensional small-rotation motion equations : uniqueness, algorithms and numerical results", CVGIP 26, 1984, pp. 183-206.

[Faug87]
O.D. Faugeras, F. Lustman and G. Toscani, "Motion and structure from motion from point and line matches", To be published, 1987.

[Hild83]
E.C. Hildreth, *The measurement of visual motion*, MIT press, Cambridge, 1983.

[Heeg86]
D.J. Heeger, "Depth and flow from motion energy", Science 86, pp. 657-663.

[Horn81]
B.K.P. Horn and B.G. Schunck, "Determining optical flow", Artificial Intelligence 17, 1981, pp. 185-204.

[Huan81]
T.S. Huang and R.Y. Tsai, "Image sequence analysis: Motion estimation", in Image Sequence Processing and Dynamic Scene Analysis, T.S. Huang, Ed. NY: Springer-verlag, 1981.

[Iu89a]
S.-L. Iu and K. Wohn, "Estimation of 3-D motion and structure based on a temporally-oriented approach with the method of regression", IEEE Workshop on visual motion, March 1989, pp. 273-281.

[Iu89b]
S.-L. Iu and K. Wohn, "Recovery of 3-D motion of a single particle", SPIE Proceedings, Intelligent Robots and Computer Vision VIII, Philadelphia, Nov. 1989, pp. 746-757.

[Iu90a]
S.-L. Iu, "Analysis of the effects of model mismatch and FLAT MMF for estimating particle motion", Tech. Report MS-CIS-90-10, GRASP Lab., Univ. of Pennsylvania, Feb. 1990.

[Kuma89]
R.V.R. Kumar, A. Tirumalai and R. C. Jain, "A non-linear optimization algorithm for the estimation of structure and motion parameters", IEEE Workshop on visual motion, March 1989, pp. 136-143.

[Kana85]
K. Kanatani, "Structure from motion without correspondence : general principle", 9th IJCAI, pp. 886-888.

[Liu86]
Y. Liu and T.S. Huang, "Estimation of rigid body motion using straight line correspondences: further results", IEEE ICPR 1986, pp. 306-309.

[Lin86]
Z.C. Lin, H. Lee and T.S. Huang, "Finding 3-D point correspondences in motion estimation", IEEE ICPR 1986, pp. 303-305.

[Long80]
H.C. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image", Proc. of Royal Society of London, B208, pp. 385-397.

[Long81]
H.C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections", Nature 293, pp. 133-135, 1981.

[Long84]
H.C. Longuet-Higgins, "The visual ambiguity of a moving plane ", Proc. of Royal Society of London, B223, pp. 165-175.

[Mayb82]
P. S. Maybeck, *Stochastic models, estimation and control*, vol. 1-2, Academic Press, 1982.

[Miti86]
A. Mitiche, S. Seida and J.K. Aggarwal, "Line-based computation of structure and motion using angular invariance", Proc. of IEEE Workshop on visual motion : representation and analysis, 1986, pp. 175-180.

[Nage81]
H.-H. Nagel, " Representation of moving rigid objects based on visual observations", Computer, Aug. 1981, pp. 29-39.

[Nage83]
H.-H. Nagel, "Displacement vectors derived from second-order intensity variations in image sequences", CVGIP, 1983, pp.85-117.

[Nage86]
H.-H. Nagel, "Image sequences-Ten (octal) years- From phenomenology towards a theoretical foundation", in Proc. Int. Conf. on Pattern Recognition, Oct. 1986, pp. 1174-1185.

[Rana80]
S. Ranade and A. Rosenfeld, "Point pattern matching by relaxation", Pattern Recognition, vol. 12, 1980, pp. 269-275.

[Roac80]
J.W. Roach and J.K. Aggarwal, "Determining the movement of object from a sequence of images", IEEE PAMI, Nov 1980, pp.554-562.

[Schu85]
B.G. Schunck, "Image flow: Fundamentals and future research", in Proc. of IEEE Conf. on Pattern recognition and Image Processing, 1985, pp. 560-571.

[Seth87]
S.K. Sethi and R. Jain, "Finding trajectories of feature points in a monocular image sequence", IEEE PAMI, Jan. 1987, pp. 56-73.

[Subb85]
M. Subbarao and A. M. Waxman, "On the uniqueness of image flow solutions for planar surfaces in motion", IEEE Workshop on computer vision: representation and control, Oct 1985, pp. 129-140.

[Tsai83]
R. Y. Tsai, "3-D inference from the motion parallax of a conic arc and a point in two perspective view", IEEE IJCAI83, pp. 1038-1042.

[Tsai84]
R.Y. Tsai and T.S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces", IEEE PAMI, Jan 1984, pp. 13-26.

[Ullm79]
S. Ullman, *The interpretation of visual motion*, Cambridge MIT Press.

[Ullm81]
S. Ullman, "Analysis of visual motion by biological and computer systems", IEEE Computer, Aug 1981, pp. 57-69.

[Waxm85]
A. M. Waxman and S. Ullman, "Surface structure and 3-D motion from image flow : a kinematic analysis", Intl. Journal of Robotics Research 4, 1985, pp. 72-94.

[Waxm86]
A.M. Waxman, B.Kamgar-Parsi and M. Subbarao, "Closed-form solutions to image flow equations for 3-D structure and motion", CAR-TR-190, Univ. of Maryland, Feb 1986.

[Weng87]
J. Weng, T.S. Huang and N. Ahuja, "3-D motion estimation, understanding and prediction from noisy image sequences", IEEE PAMI, May 1987, pp. 370-389.

[Wohn83]
K. Wohn, L.S. Davis and P. Thrift, "Motion estimation based on multiple local constraints and nonlinear smoothing", Pattern Recognition 16, 1983, pp. 563-570.

[Wu86]
J. Wu and K. Wohn, "Recovering 3-D motion and structure from 1st-order image deformation", SPIE symposium on intelligent robots, Oct 1986.

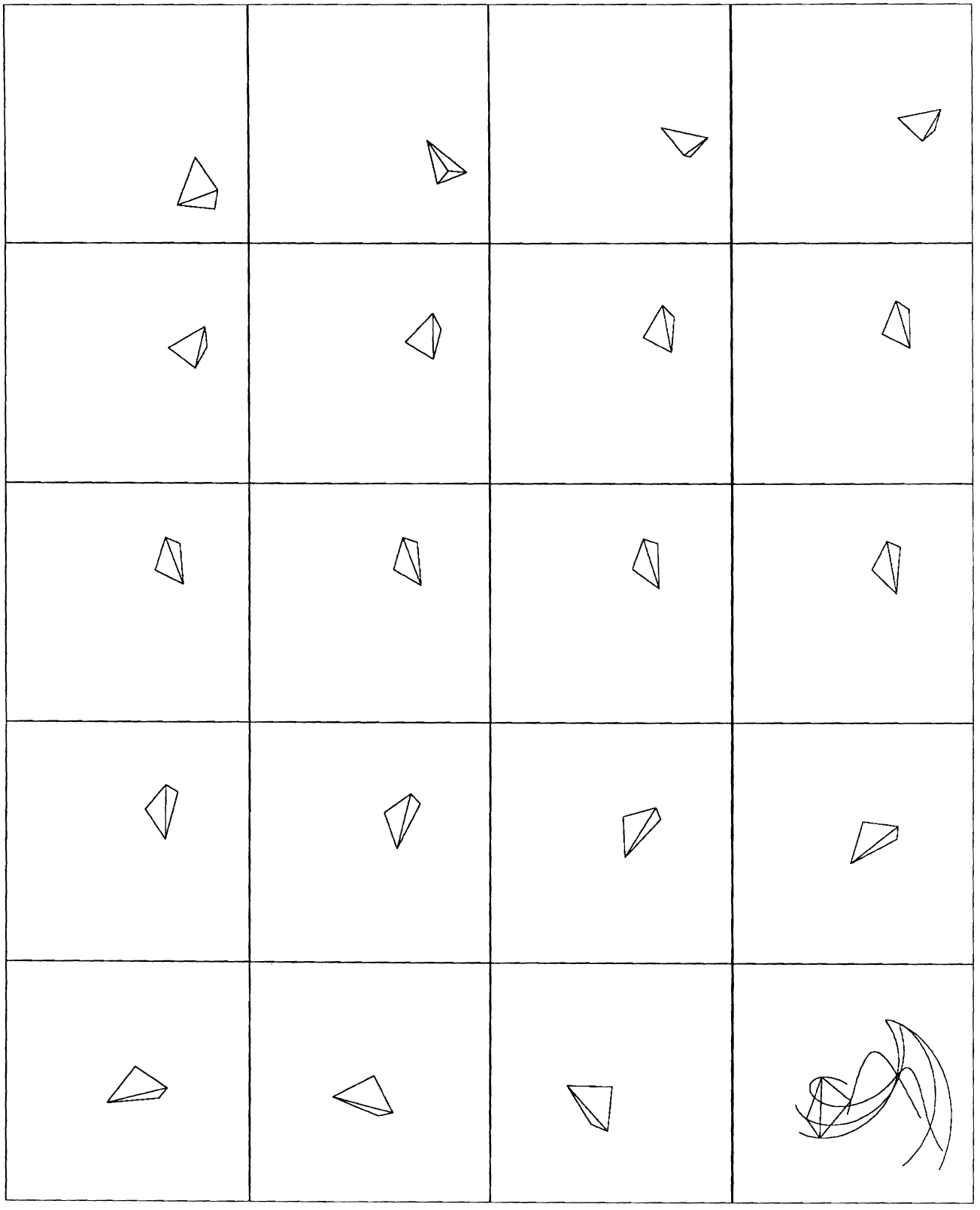

Fig. 1a. Images of simulated motion at every five frames for experiment 1. Last plot includes the trajectories on image plane.

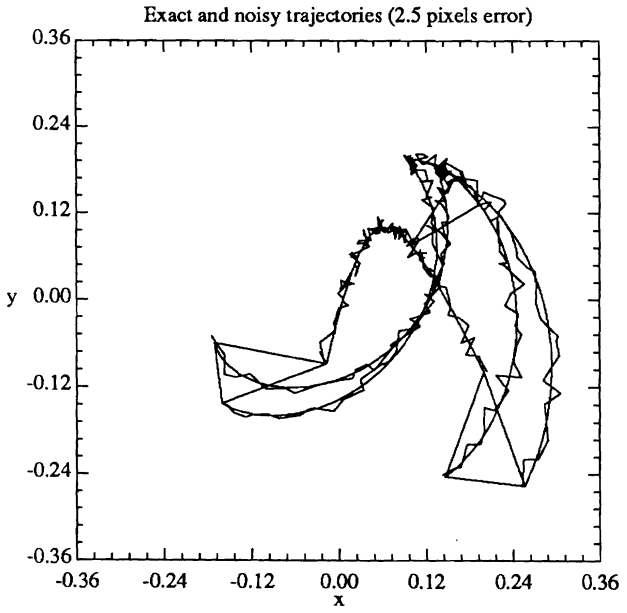

Fig. 1b. Exact and noisy trajectories for experiment 1. Standard deviation of the noise is 2.5 pixels.

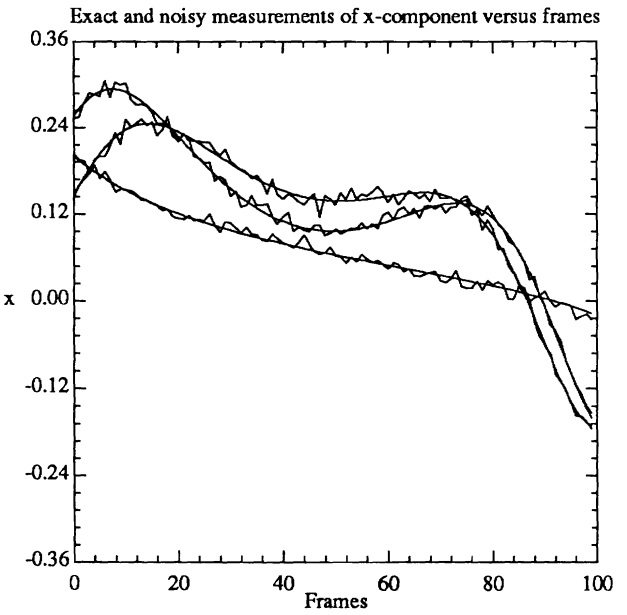

Fig. 1c. Noisy measurements of x-component of the trajectories versus number of frames for experiment 1. Standard deviation of the noise is 2.5 pixels.

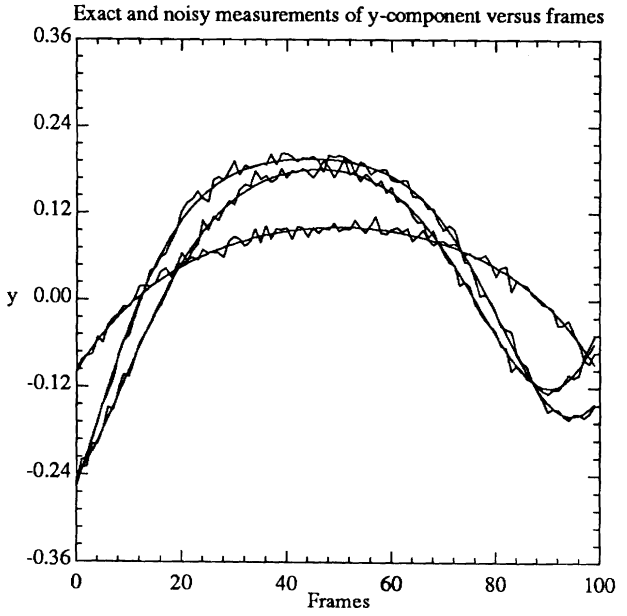

Fig. 1d. Noisy measurements of y-component of the trajectories versus number of frames for experiment 1. Standard deviation of the noise is 2.5 pixels.

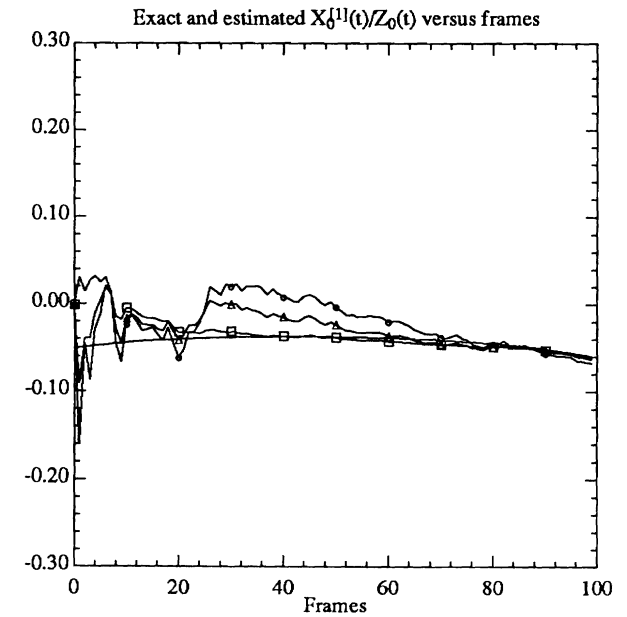

Fig. 1e. Exact and estimated $X_0^{[1]}(t)/Z_0(t)$ versus number of frames for experiment 1. Standard deviations of the noise are 0.5 (□), 2.5 (Δ) and 5.0 (O) pixels.

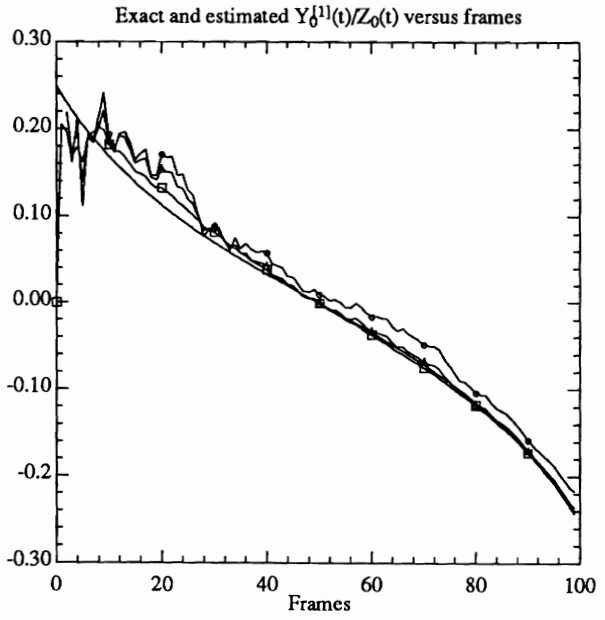

Fig. 1f. Exact and estimated $Y_0^{[1]}(t)/Z_0(t)$ versus number of frames for experiment 1. Standard deviations of the noise are 0.5 (□), 2.5 (Δ) and 5.0 (O) pixels.

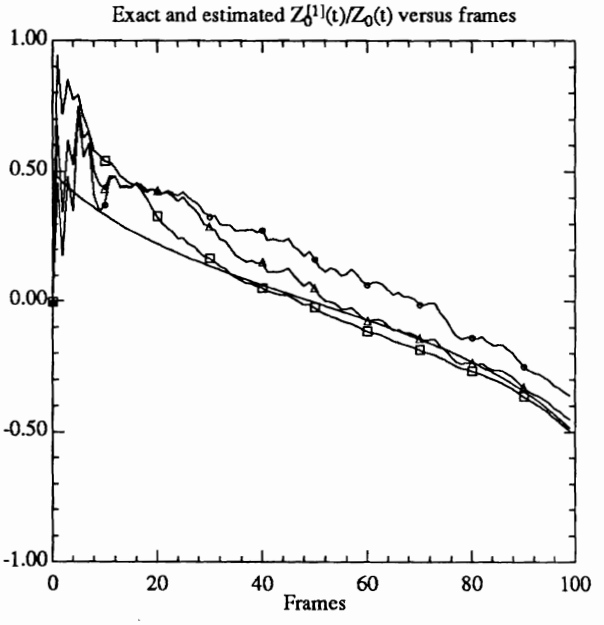

Fig. 1g. Exact and estimated $Z_0^{[1]}(t)/Z_0(t)$ versus number of frames for experiment 1. Standard deviations of the noise are 0.5 (□), 2.5 (Δ) and 5.0 (O) pixels.

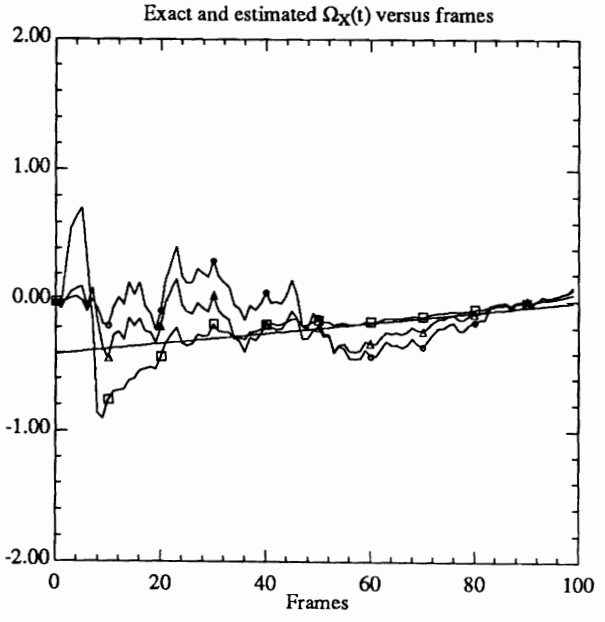

Fig. 1h. Exact and estimated $\Omega_X(t)$ versus number of frames for experiment 1. Standard deviations of the noise are 0.5 (□), 2.5 (Δ) and 5.0 (O) pixels.

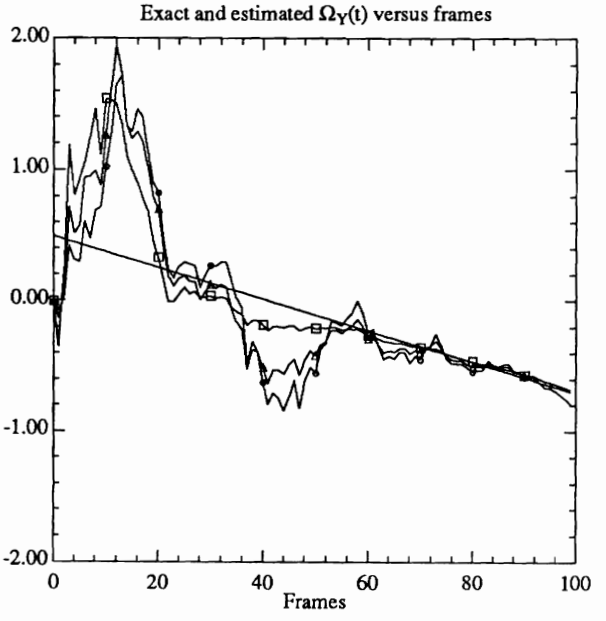

Fig. 1i. Exact and estimated $\Omega_Y(t)$ versus number of frames for experiment 1. Standard deviations of the noise are 0.5 (□), 2.5 (Δ) and 5.0 (O) pixels.

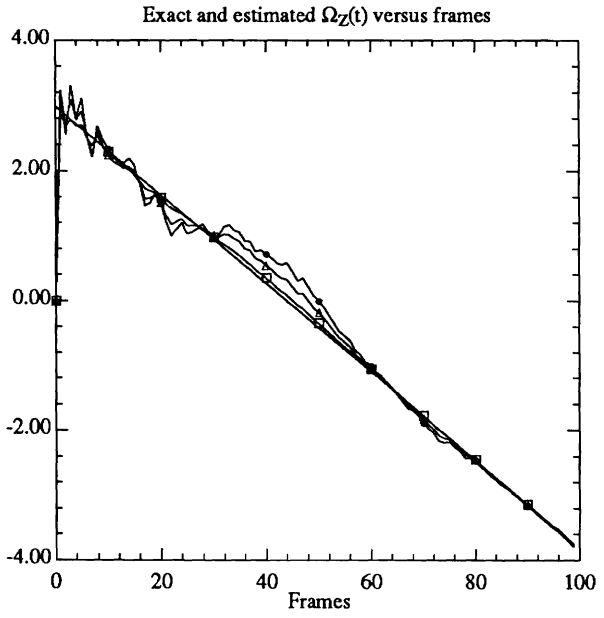

Fig. 1j. Exact and estimated $\Omega_Z(t)$ versus number of frames for experiment 1. Standard deviations of the noise are 0.5 (□), 2.5 (Δ) and 5.0 (○) pixels.

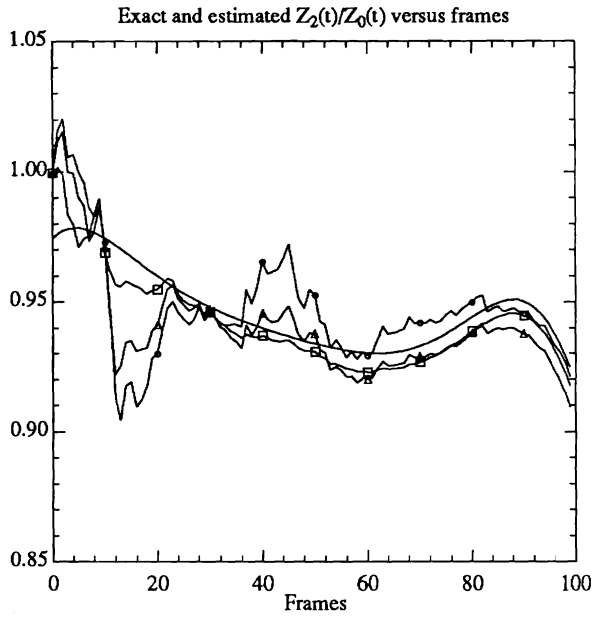

Fig. 1k. Exact and estimated $Z_1(t)/Z_0(t)$ versus number of frames for experiment 1. Standard deviations of the noise are 0.5 (□), 2.5 (Δ) and 5.0 (○) pixels.

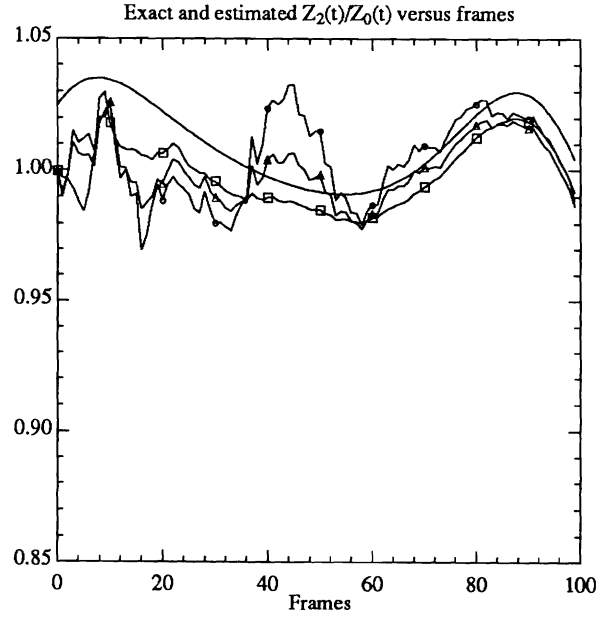

Fig. 1l. Exact and estimated $Z_2(t)/Z_0(t)$ versus number of frames for experiment 1. Standard deviations of the noise are 0.5 (□), 2.5 (Δ) and 5.0 (○) pixels.

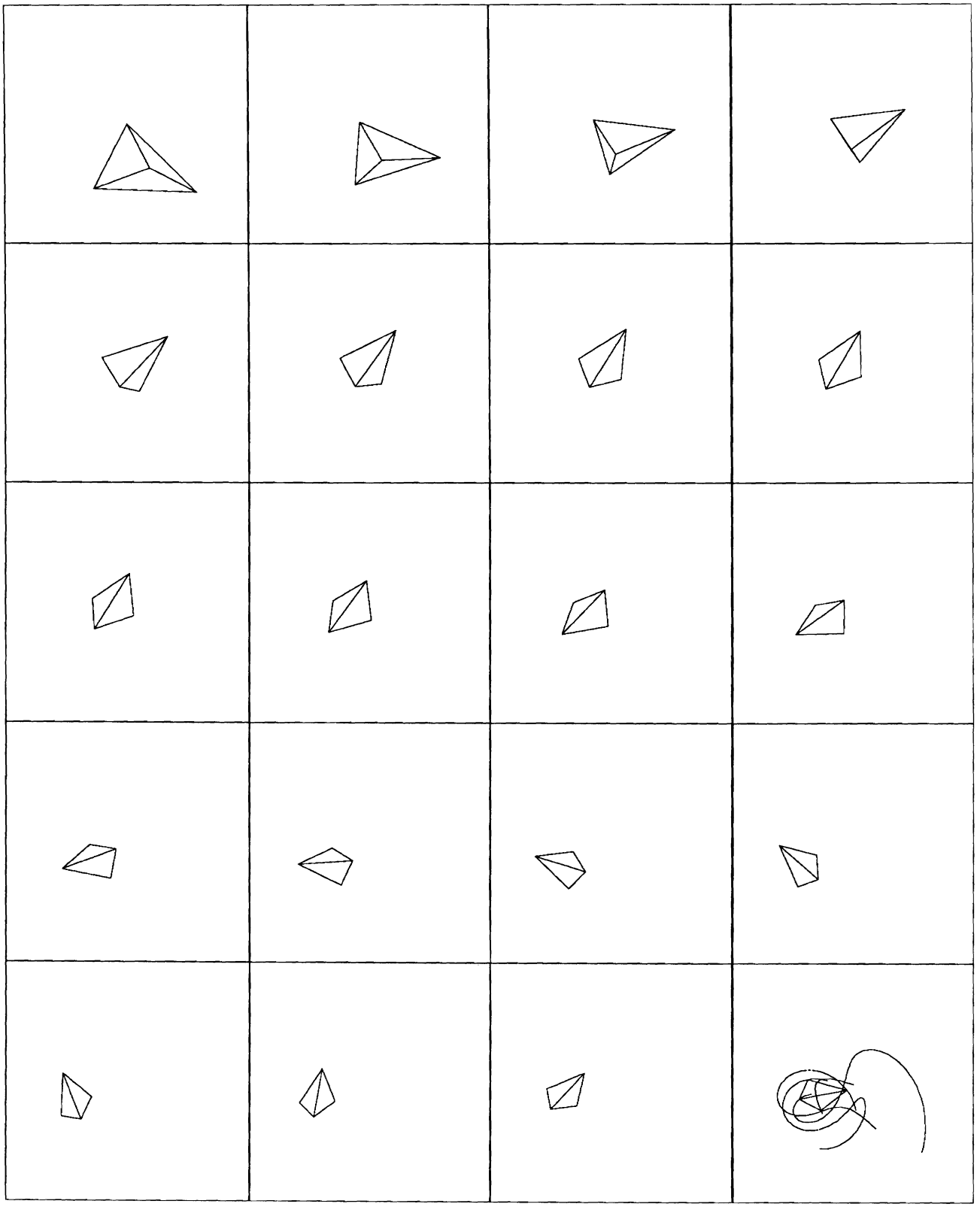

Fig. 2a. Images of simulated motion at every five frames for experiment 2. Last plot includes the trajectories on image plane.

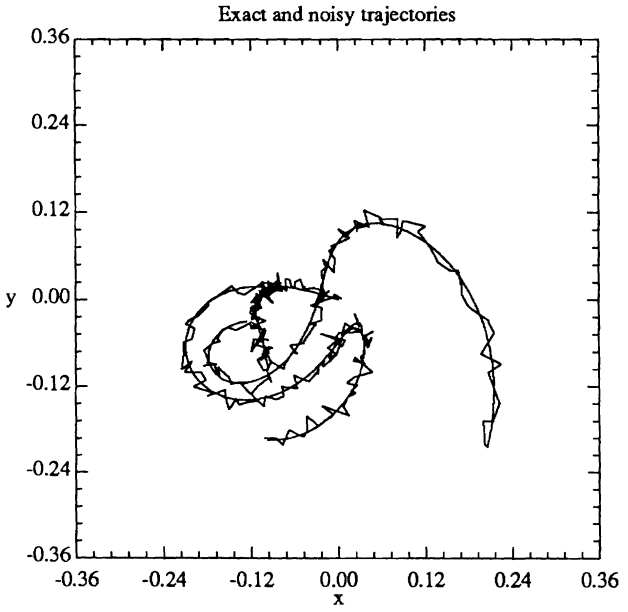

Fig. 2b. Exact and noisy trajectories for experiment 2. Standard deviation of the noise is 2.5 pixels.

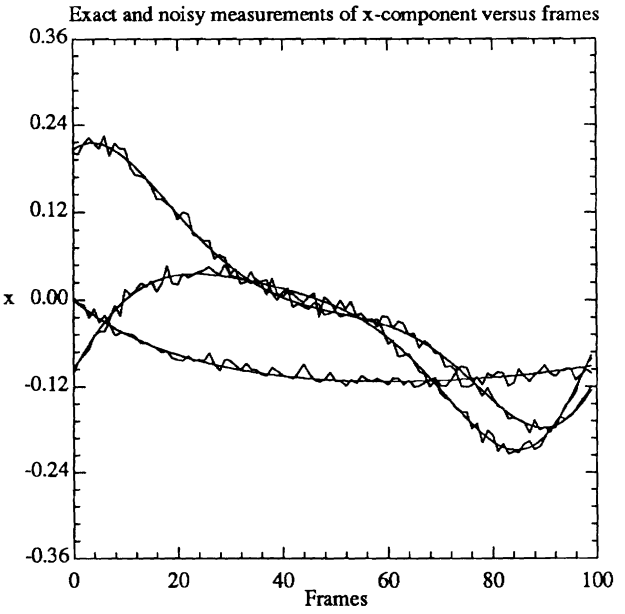

Fig. 2c. Noisy measurements of x-component of the trajectories versus number of frames for experiment 2. Standard deviation of the noise is 2.5 pixels.

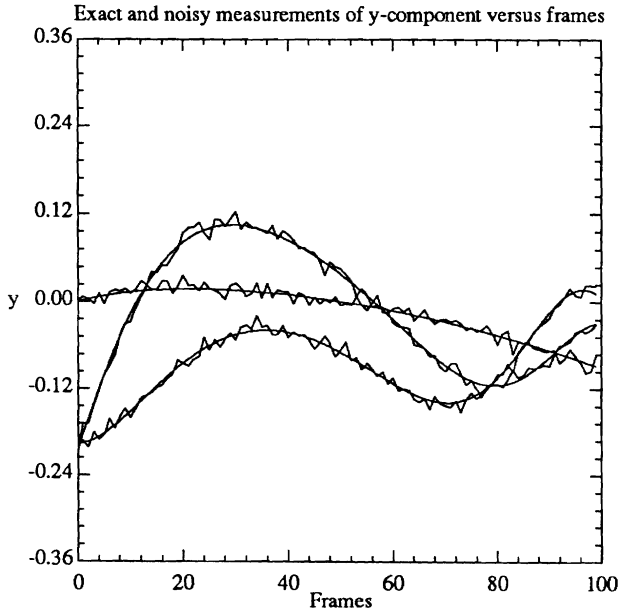

Fig. 2d. Noisy measurements of y-component of the trajectories versus number of frames for experiment 2. Standard deviation of the noise is 2.5 pixels.

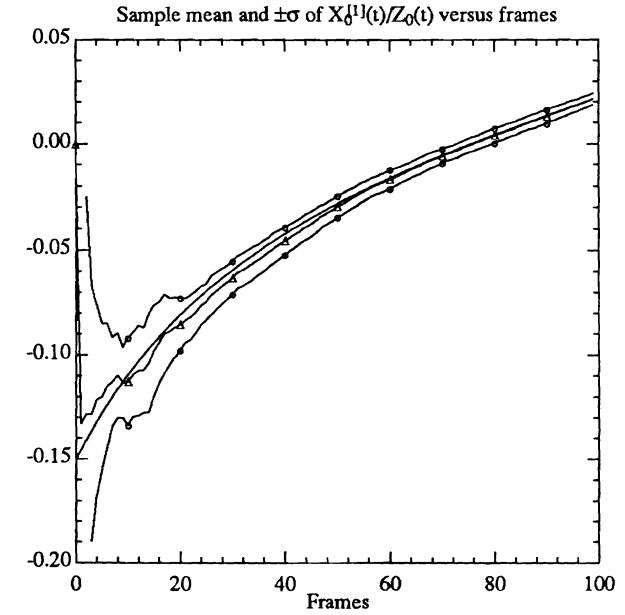

Fig. 2e. Sample mean (Δ) and ± standard deviation (O) of estimated $X_0^{[1]}(t)/Z_0(t)$ versus number of frames for experiment 2.

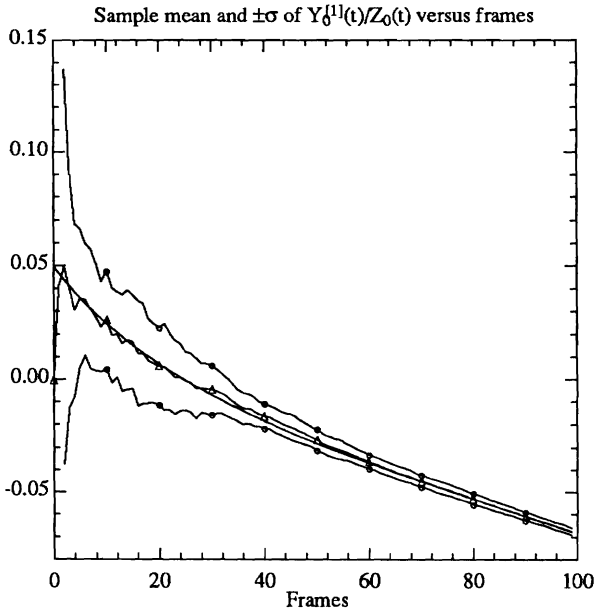

Fig. 2f. Sample mean (Δ) and ± standard deviation (O) of estimated $Y_\delta^{[1]}(t)/Z_0(t)$ versus number of frames for experiment 2.

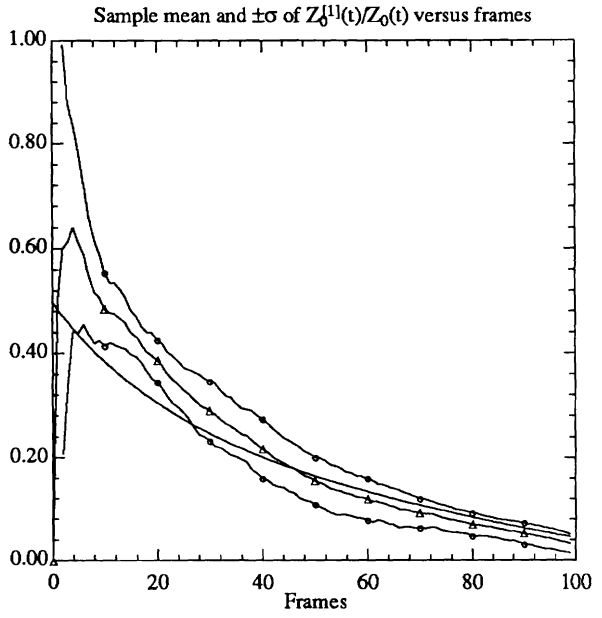

Fig. 2g. Sample mean (Δ) and ± standard deviation (O) of estimated $Z_\delta^{[1]}(t)/Z_0(t)$ versus number of frames for experiment 2.

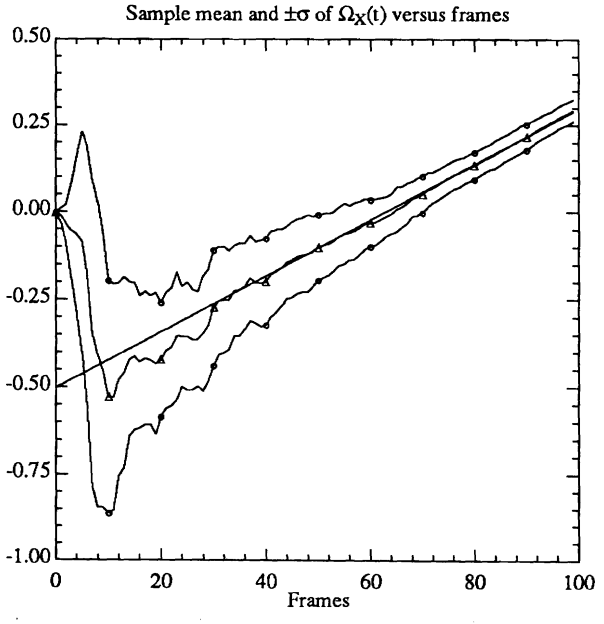

Fig. 2h. Sample mean (Δ) and ± standard deviation (O) of estimated $\Omega_X(t)$ versus number of frames for experiment 2.

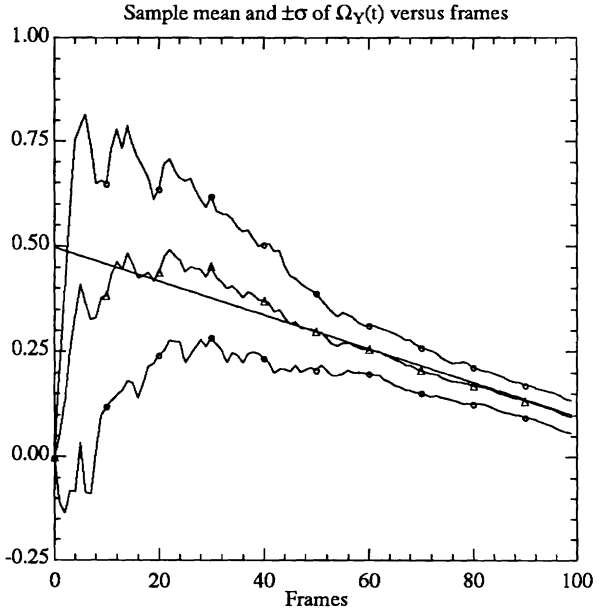

Fig. 2i. Sample mean (Δ) and ± standard deviation (O) of estimated $\Omega_Y(t)$ versus number of frames for experiment 2.

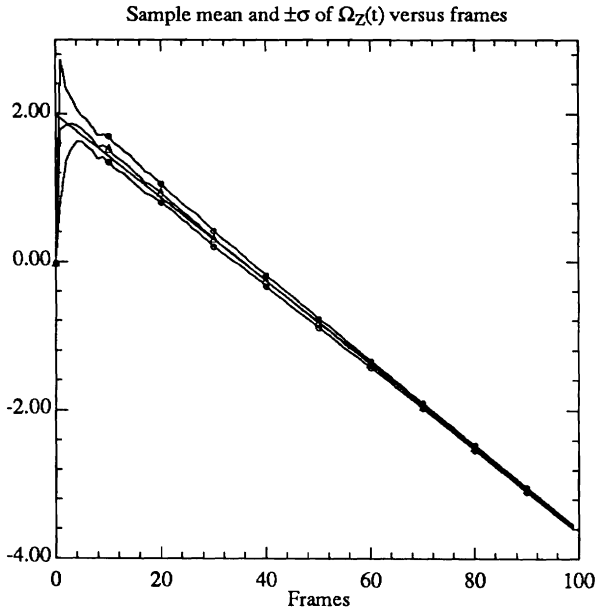

Fig. 2j. Sample mean (Δ) and ± standard deviation (O) of estimated $\Omega_Z(t)$ versus number of frames for experiment 2.

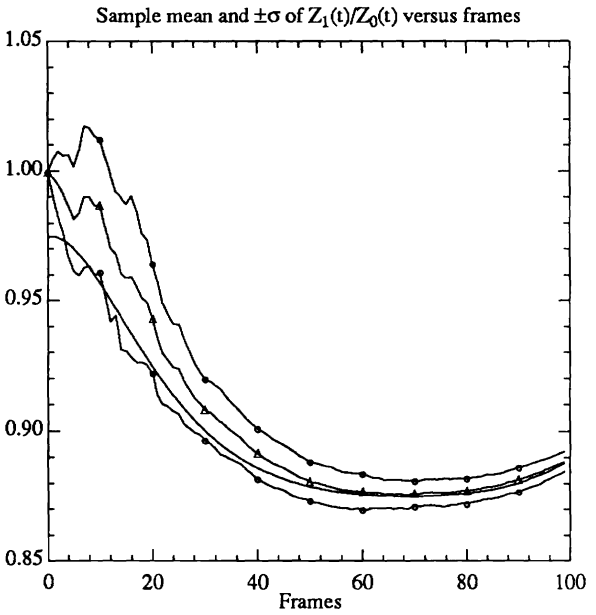

Fig. 2k. Sample mean (Δ) and ± standard deviation (O) of estimated $Z_1(t)/Z_0(t)$ versus number of frames for experiment 2.

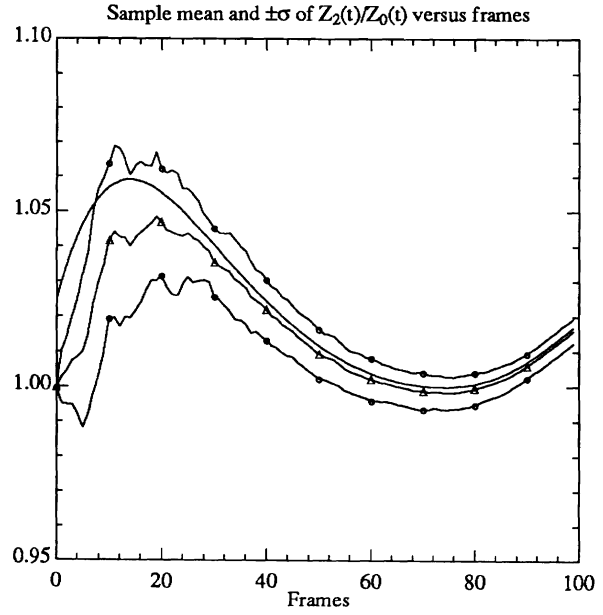

Fig. 2l. Sample mean (Δ) and ± standard deviation (O) of estimated $Z_2(t)/Z_0(t)$ versus number of frames for experiment 2.