November 1988

Estimation of 3-D Motion and Structure Based on a Temporally-Oriented Approach With the Method of Regression

Siu-Leong Iu
University of Pennsylvania

Kwangyoen Wohn
University of Pennsylvania

Estimation of 3-D Motion and Structure Based on a Temporally-Oriented Approach With the Method of Regression

Abstract

In this paper we argue that the 3-D velocity of a single point up to a scalar factor can be recovered from its 2-D trajectory under the perspective projection. We then extend the idea to the recovery of 3-D motion of rigid objects. In both cases measurements are collected through temporal axis first, while keeping the amount of measurements in each frame minimal. We may use multiple features to get a more accurate estimate if they are available. This approach called *temporally oriented approach* requires us to introduce the explicit model for the evolution of 3-D motion. Our analysis is based on the assumption that the 3-D motion is smooth so that its 3-D velocity can be approximated as a truncated Taylor series. Regression relations between unknown motion parameters and measurements for a single point and rigid body are derived. The method of Maximum Likelihood is used to estimate the motion. The uniqueness of determining the 3-D motion of a single point is discussed. Experimental results obtained from simulated data and real images are given to illustrate the robustness of this approach.

Comments

ESTIMATION OF 3-D MOTION
AND STRUCTURE BASED ON A
TEMPORALLY-ORIENTED APPROACH
WITH THE METHOD OF REGRESSION

SiuLeong Iu
and K. Wohn

MS-CIS-88-95
GRASP LAB 167

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

November 1988

ESTIMATION OF 3-D MOTION AND STRUCTURE BASED ON A TEMPORALLY-ORIENTED APPROACH WITH THE METHOD OF REGRESSION

Abstract

In this paper we argue that the 3-D velocity of a single point up to a scalar factor can be recovered from its 2-D trajectory under the perspective projection. We then extend the idea to the recovery of 3-D motion of rigid objects. In both cases measurements are collected through temporal axis first, while keeping the amount of measurements in each frame minimal. We may use muliple features to get a more accurate estimate if they are available. This approach called *temporally oriented approach* requires us to introduce the explicit model for the evolution of 3-D motion. Our analysis is based on the assumption that the 3-D motion is smooth so that its 3-D velocity can be approximated as a truncated Taylor series. Regression relations between unknown motion parameters and measurements for a single point and rigid body are derived. The method of Maximum Likelihood is used to estimate the motion. The uniqueness of determining the 3-D motion of a single point is discussed. Experimental results obtained from simulated data and real images are given to illustrate the robustness of this approach.

1. INTRODUCTION

Estimating the 3-D motion and structure of objects in space from image is one of the major problems in computer vision. Resulting estimates may play an important role in building intelligent robot, tracking moving objects and autonomous navigation. In the past few years, a substantial amount of work have been done by many researchers in finding these estimates. However, it is widely held that most existing schemes are very sensitive to noise even when synthetically generated measurements are used [Agga85, Fang84, Wu86, Broi86b, Roac80, Lin86]. In this paper, a new algorithm which relies on a large number of images is proposed. Many experimental results dealt with simulation measurements and real images are given to illustrate the robustness of this approach.

There are two kinds of measurements we can obtain from image squences. They are, respectively, the measurements of projective positions [Agga81, Lin86, Seth87, Fang84, Rana80], which are defined as the positions in the image coordinates corresponding to 3-D feature points on the object, and the measurements of optical flows [Ullm81, Horn81, Wohn83, Hild83, Nage83, Schu85, Hara81, Tret84, Enke84, Heeg86]. Using these two measurements lead to two different kinds of approaches in recovering the 3-D motion: token-matching approaches and optical flow approaches.

For the token matching approaches, point features corresponding to motion of points in the scene have been studied extensively [Ullm79, Roac80, Nage81, Long81, Huan81, Tsai81, Tsai82, Tsai84, Zhuan86, Faug87, Nage86]. Other features like straight lines in the scene have also been used. Liu and Huang [Liu86] showed that the rigid body motion can be determined by using six lines over three frames. Mitiche, Seida and Aggarwal [Miti86] used the conservation of angular configuration as a rigidity constraint and showed that four lines in three frames are sufficient to determine the rigid body motion. Faugeras, Lustman and Toscani [Faug87] showed that at least three views are necessary when line segments are used. Tsai studied the conic arc to determine the motion [Tsai83].

For the optical flow approach, 3-D motion is determined from the mesurements of the optical flows and their temporal and spatial derivatives. Longuet-Higgins and Prazdny [Long80] showed that the rigid motion and local orientation of a curved surface patch could be determined by using the spatial derivatives of the optical flow up to second order. Waxman and Ullman [Waxm85] derived a kinematic approach to optical flow analysis from the fluid mechanics. They showed that the slopes, curvatures and motion could be recovered for both planar and curved surfaces. Longuet-Higgins [Long84], Subbarao and Waxman [Subb85] and Kanatani [Kana85] have independently obtained closed form solutions for planar surfaces in motion. Waxman, Kamgar-Parsi and Subbarao [Waxm86a] obtained close-form solutions for curved surfaces based on the solution of planes. Wu and Wohn

[Wu86] recovered the rigid body motion and the surface normal by using the spatial and temporal derivatives up to first order.

Aggarwal and Nandhakumar [Agga88] gave an excellent and up to date review of the whole field of estimating 3-D structure and motion from sequences of monocular and stereoscopic images. They described some of the different mathematical formulations of the token-matching approach and optical-flow approach, and compared in detail of these two different approaches. They also presented an oview of the fusion of stereo analysis and motion analysis, and discussed different algorithms in finding the correspondence between the features in the image sequences.

Although the token matching approach and optical flow approach use different measurements of the motion and follow different procedures in estimating the 3-D motion and sturcture of the object, it is observed that they are biased more toward the spatial information of image sequence than the temporal information. In the token matching approach, the main theme is to find the minimal number of points and condition that guarantee the unique solution, when N (=2,3, typically) frames are given. Similarly, the optical flow approach requires spatial derivatives of the optical flow field evaluated from two or three frames. Here, the theme is "Find the minimal order spatial derivatives when the flow field is given." Of course, these two approaches do use the temporal change of the images, but they use the images separately. Most of them use only two or three frames at a time. Let us call this kind of approach a *spatially-oriented approach* (SOA).

There are a number of problems we need to face when the SOAs are persued. Firstly, if the moving object does not have enough features on it, then SOAs can not estimate the 3-D motion of this object. Secondly, we need to segment the scene such that we can use the features belonging to the same object. The more features we use, the deeper level of segmentation required. Thirdly, if we only use two or a few frames to find the 3-D motion, no matter what method we use, the estimation problem becomes ill-posed itself. Even for human beings, we have difficulty even in making qualitative interpretation for the 3-D motion when we have only two snapshots of a moving object. If the time interval between these two snapshots is very small, we may assume the motion is constant between the snapshots and use them to estimate the motion. But the small change in these snapshots make it very difficult to determine the motion, especially when there are noises in the observation. If the time interval between the snapshots is large, we can not even "imagine" how the object moves between our observations.

In this paper, we propose a new approach to the estimation of 3-D motion from image sequences. This approach seeks for the temporal information prior to the spatial information, such as the trajectory

of a moving point in the image plane or the temporal change of the slope of a line which is the projection of a 3-D line. We call this kind of approach a *temporally-oriented approach* (TOA). The importance of TOA is that we can avoid many problems we face in SOAs. Since we observe motion over an extended period of time, we can reduce the number of features which are required by SOAs. Actually, we will show that we can analyze the motion even for a single particle. Consequently, the problem of requiring multiple features can be eliminated and the burden for the segmentation is reduced. Also, as we use more frames to estimate the 3-D motion, the problem itself becomes more well-posed. This can be justified as follows. When we observe a moving object, such as a flying shuttle and a hunting leopard, the longer we observe, the more accurately we can estimate its motion and predict where this object will be.

A number of researchers have proposed the use of a large number of frames in estimating 3-D motion in order to improve estimates and/or to simplify the procedure in finding the measurements of motion [Weng87, Broi86a,b, Wu86, Waxm86b, Boll85]. However, their approaches just reverse the thinking of TOA. Their motivation for using multiple frames is that they want to use temporal information to "help" or "stabilize" spatially-oriented approaches in determining the motion. In TOAs, we rely on the temporal information from the moving object while keeping the number of feature points in a single image as small as possible. Of course we may use multiple features to get a more robust estimate if they are available. Therefore, the TOA and the previous multi-frame approach are different in their motivation. The following is some reviews of the previous work which utilizes a large number of frames.

Weng, Huang and Ahuja [Weng87] proposed the locally constant angular mometum model in estimating 3-D motion from the measurements of projective positions. Their model assumed that angular momentum is constant over short time intervals, the moving body possesses an axis of symmetry and the motion of the rotation center is approximated by 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 derive directly the relation between the unknown motion parameters and the projective positions, rather than solving the equations for the external torque and the angular momentum. In finding the 3-D motion, they first estimated the rotation matrices and translations between the frames by using "two-frame" motion analysis and then estimated the motion of the rotation center from these estimated rotation matrices and translation, i.e., their approach still considers the frames 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.

Wu and Wohn [Wu86] recovered 3-D rigid body motion and the surface normal from optical flows by using the spatial and temporal derivatives of optical flows up to the first order. The estimate was more robust than that of the previous methods which require spatial derivatives up to the second order. In finding 3-D motion, their approach uses more temporal information but still relies on spatial information.

Broida and Chellappa [Broi86a,b] estimated motion parameters sequentially by using projective positions of multiple points in a sequence of noisy images. They used the dynamic model to describe the temporal behavior of the estimated parameters and the iterative extended Kalman filter to find the estimate. In [Broi86a], they estimated the motion of a two-dimensional object by using one-dimensional projected images. They assumed that the image coordinates of the 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 the estimation of 3-D rigid body motion. Motion was modeled by a truncated Taylor series but only the linear terms were used in their derivation. The simulation was conducted only on very special motion such as pure translation or rotation about a fixed and known axis. The basic difference between their work and our approach is that they use the nonlinear, coupled vector differential equation to describe the relationship between the unknown parameters at different time while we use the regression relation to describe this relationship explicitly. The number of the unknown parameters they needed to estimate are more than three times of that of our approach.

Bolles and Baker [Boll85] proposed the "Epipolor-plane image" (EPI) analysis to estimate the pure translation of a camera or the absolute depth in a stationary enviroment. They used very dense image sequences in estimation so that the estimates could be determined by simply measuring the slopes of the lines on EPI planes. Hence, their motivation for using multiple frames was to simplify the estimation procedure. They pointed out that their geometric approach does not work for general motion.

The key step in our temporally-oriented approach is to integrate the information from each frame effectively so that we can use the measurements in all available frames. We will derive the regression relations [Spre69, Will59] for the motion of particle and rigid body in which the unknown motion parameters are related to the projective positions explicitly. Then we will use the method of Maximum Likelihood to estimate the motion from noisy measurements of the projective position. Note that in using these regression relations to find the estimate, we will include all the measurements for different

feature points on the rigid body. This means that we use all the information in the temporal domain as well as the spatial domain in estimating the motion. Furthermore, we can use the previous estimates of the parameters, as many regression models do, to predict where the next projective positions will be. This is important in obtaining the reliable measurement of the projected positions because this prediction will decrease the chance of mis-matching the features and reduce the size of the searching space of the moving features on the image plane. Also, this prediction is useful in solving the problem of object occlusion.

In this paper, we assume that the measurements of projective positions corrupted with noise are available. We assume that the motion of the object is smooth so that the 3-D velocity of the motion can be approximated as the truncated Taylor series. Figure 1 shows the camera model and the coordinate system we will use in our derviation.

The rest of this paper is organized as follows. Section 2 studies the motion of a particle. We prove that we can determine uniquely the 3-D velocity of the motion up to a scale factor from the projective position sequence. We then derive the regression relation between motion parameters and projective positions. The method of Maximum Likelihood is used to estimate the motion from noisy measurements. Experimental results and discussions follow. Section 3 studies the motion of rigid body. We propose to use the state equation approach to find the regression relations between unknown motion parameters and projective positions of the rigid body. The state equation describes the evolution of motion in time. We first consider the case of constant translation and rotation and then extend the study to general motion. We also investigate how to reduce the number of unknowns if all the points lie on a planar surface. The method of Maximum Likelihood is used to find the estimate. Simulation results and discussion are given. Section 4 is the conclusion of this paper.

2. MOTION ANALYSIS OF A MOVING PARTICLE

In this section, we will estimate the instantaneous 3-D motion of a moving particle by using the sequence of its projective position. Section 2.1 studies the model of the motion and the uniqueness of the estimation. Section 2.2 discusses how to estimate the motion parameters when measurements are noisy. Experimental results and discussions are given in section 2.3.

2.1. MODEL OF THE MOTION AND UNIQUENESS OF THE SOLUTION

Let $\underline{R}(t) = [X(t)\ Y(t)\ Z(t)]^T$, $[x(t)\ y(t)]^T$ and $\underline{V}(t) = [V_X(t)\ V_Y(t)\ V_Z(t)]^T$ be the 3-D position, the projective position and the instantaneous velocity of a moving particle at time t, respectively. Let $\underline{V}'(t) = [V'_X(t)\ V'_Y(t)\ V'_Z(t)]^T$ be the velocity $\underline{V}(t)$ scaled by $1/Z(0)$. Then we have

$$\frac{d}{dt}\underline{R}(t) = \underline{V}(t)$$ (2.1.1)

$$x(t) = X(t) / Z(t), \quad y(t) = Y(t) / Z(t) \qquad\qquad (2.1.2a,b)$$

If the particle moves smoothly in the 3-D space, we may model the $\underline{V}(t)$ as a truncated Taylor series and express the components of $\underline{V}(t)$ in the following form.

$$V_X(t) = \sum_{i=0}^{n_X} V_{Xi}\, t^i, \quad V_Y(t) = \sum_{i=0}^{n_Y} V_{Yi}\, t^i, \quad V_Z(t) = \sum_{i=0}^{n_Z} V_{Zi}\, t^i \qquad\qquad (2.1.3a,b,c)$$

Note that the coefficients in the above expansion have their physical interpretation. For example, $[V_{X0}\ V_{Y0}\ V_{Z0}]^T$ and $[V_{X1}\ V_{Y1}\ V_{Z1}]^T$ represent the velocity and the acceleration of the moving particle at time zero, respectively. The selection of n_X, n_Y and n_Z in (2.1.3), the order of the motion, is dependent on the temporal behavior of the motion, i.e., how smooth the object moves in the 3-D space. It is well known that we can only estimate the translation of a moving object up to a scale factor because of the perspective projection. Thus, our objective is to find the unknown coefficients in (2.1.3) scaled by $1/Z(0)$ from the measurements of the sequence of the projective position $[x(t)\ y(t)]^T$.

Integrating (2.1.1) and using (2.1.3), we obtain

$$X(t) = X(0) + \sum_{i=0}^{n_X} V_{Xi}\, \frac{t^{i+1}}{i+1}, \quad Y(t) = Y(0) + \sum_{i=0}^{n_Y} V_{Yi}\, \frac{t^{i+1}}{i+1}, \quad Z(t) = Z(0) + \sum_{i=0}^{n_Z} V_{Zi}\, \frac{t^{i+1}}{i+1} \qquad (2.1.4a,b,c)$$

Consequently, we have

$$x(t) = \frac{x(0) + \sum_{i=1}^{n_X+1} V'_{X(i-1)}\, \dfrac{t^i}{i}}{1 + \sum_{i=1}^{n_Z+1} V'_{Z(i-1)}\, \dfrac{t^i}{i}}, \qquad y(t) = \frac{y(0) + \sum_{i=1}^{n_Y+1} V'_{Y(i-1)}\, \dfrac{t^i}{i}}{1 + \sum_{i=1}^{n_Z+1} V'_{Z(i-1)}\, \dfrac{t^i}{i}} \qquad\qquad (2.1.5a,b)$$

where $\{V'_{Xi}, i = 0, ..., n_X\}$, $\{V'_{Yi}, i = 0, ..., n_Y\}$ and $\{V'_{Zi}, i = 0, ..., n_Z\}$ are the coefficients in (2.1.3) scaled by $1/Z(0)$.

If $N(t)$ and $D(t)$ are polynomials with degrees n_N and n_D respectively and the constant coefficient of $D(t)$ is one, then $N(t)/D(t)$ is referred to as a rational function with degree (n_N, n_D). Thus, from (2.1.5), the components of the projective position $x(t)$ and $y(t)$ are rational functions with degrees $(n_X + 1, n_Z + 1)$ and $(n_Y + 1, n_Z + 1)$, respectively.

If we know all the coefficients in (2.1.5), then the trajectory $[x(t)\ y(t)]^T$ on the image plane over the time is uniquely determined. Our motion analysis just reverses this correspondence, i.e., we want to determine these coefficients from the given trajectory. It is worth asking whether or not this determination is unique when the orders of the model of the velocity in (2.1.3) have been prefixed. In the following we will show that, from the projective positions of a moving particle, we can recover the instantaneous velocity up to a scale factor uniquely. First, two related lemmas are given.

Lemma 1

If $x(t)$ and $y(t)$ are irreducible rational functions on $t \in [t_1, t_2]$ with degrees $(n_X + 1, n_Z + 1)$ and $(n_Y + 1, n_Z + 1)$, respectively, and $x(t)$ and $y(t)$ are not equal to zero, then the 1×2 functions $\underline{f}_i(t)$, $i = 1, \dots, n_X + n_Y + n_Z + 3$ defined as

$$\underline{f}_i(t) = \begin{cases} \left[\quad \dfrac{t^i}{i} \qquad\qquad 0 \qquad\quad \right], & i = 1, \dots, n_X + 1 \\[3mm] \left[\quad 0 \qquad\qquad \dfrac{t^{i-n_X-1}}{i-n_X-1} \quad \right], & i = n_X + 2, \dots, n_X + n_Y + 2 \\[3mm] \left[-x(t)\dfrac{t^{i-n_X-n_Y-2}}{i-n_X-n_Y-2} \quad -y(t)\dfrac{t^{i-n_X-n_Y-2}}{i-n_X-n_Y-2} \right], & i = n_X + n_Y + 3, \dots, n_X + n_Y + n_Z + 3 \end{cases} \qquad (2.1.6)$$

are linearly independent on $[t_1, t_2]$ over the field of real numbers.

Proof: We prove this lemma by contradiction. Assume that $\underline{f}_i(t)$, $i = 1, \dots, n_X + n_Y + n_Z + 3$, are linearly dependent on $[t_1, t_2]$ over the field of real numbers, then there exist real numbers α_i, $i = 1, \dots, n_X + n_Y + n_Z + 3$, not all zero, such that

$$\sum_{i=1}^{n_X + n_Y + n_Z + 3} \alpha_i \, \underline{f}_i(t) = [0 \ 0] \qquad (2.1.7)$$

for all t in $[t_1, t_2]$. This implies,

$$\sum_{i=1}^{n_X+1} \alpha_i \frac{t^i}{i} - x(t) \sum_{i=1}^{n_Z+1} \alpha_{i+n_X+n_Y+2} \frac{t^i}{i} = 0 \qquad (2.1.8a)$$

$$\sum_{i=1}^{n_Y+1} \alpha_{i+n_X+1} \frac{t^i}{i} - y(t) \sum_{i=1}^{n_Z+1} \alpha_{i+n_X+n_Y+2} \frac{t^i}{i} = 0 \qquad (2.1.8b)$$

If $\alpha_{i+n_X+n_Y+2} = 0$, $i = 1, \dots, n_Z + 1$, then

$$\sum_{i=1}^{n_X+1} \alpha_i \frac{t^i}{i} = 0, \qquad \sum_{i=1}^{n_Y+1} \alpha_{i+n_X+1} \frac{t^i}{i} = 0 \qquad (2.1.9a,b)$$

This implies $\alpha_i = 0$ for all i, which contradicts the hypothesis that α_i's are not all zero. If $\alpha_{i+n_X+n_Y+2} \neq 0$, $i = 1, \dots, n_Z + 1$, then

$$x(t) = \frac{\displaystyle\sum_{i=1}^{n_X+1} \alpha_i \frac{t^i}{i}}{\displaystyle\sum_{i=1}^{n_Z+1} \alpha_{i+n_X+n_Y+2} \frac{t^i}{i}} = \frac{\displaystyle\sum_{i=1}^{n_X+1} \alpha_i \frac{t^{i-1}}{i}}{\displaystyle\sum_{i=1}^{n_Z+1} \alpha_{i+n_X+n_Y+2} \frac{t^{i-1}}{i}} \qquad (2.1.10a)$$

$$y(t) = \frac{\displaystyle\sum_{i=1}^{n_Y+1} \alpha_{i+n_X+1} \frac{t^i}{i}}{\displaystyle\sum_{i=1}^{n_Z+1} \alpha_{i+n_X+n_Y+2} \frac{t^i}{i}} = \frac{\displaystyle\sum_{i=1}^{n_Y+1} \alpha_{i+n_X+1} \frac{t^{i-1}}{i}}{\displaystyle\sum_{i=1}^{n_Z+1} \alpha_{i+n_X+n_Y+2} \frac{t^{i-1}}{i}} \qquad (2.1.10b)$$

Consequently, we obtain that $x(t)$ and $y(t)$ are rational functions with degrees (n_X, n_Z) and (n_Y, n_Z), respectively. This contradicts the assumption of $x(t)$ and $y(t)$ are irreducible. Hence, the $\underline{f}$'s are linearly

independent on $[t_1, t_2]$.   Q.E.D.

Lemma 2

Let $\underline{f}_i(t)$, for i=1, ..., n, be 1× p real-valued continuous functions defined on $[t_1, t_2]$. Let F be the n×p matrix with $\underline{f}_i$ as its row, i.e., $F \equiv [\underline{f}_1^T(t)\ \underline{f}_2^T(t)\ \ldots\ \underline{f}_n^T(t)]^T$. Define the Gram matrix of the $\underline{f}$'s as

$$W(t_1, t_2) \equiv \int_{t_1}^{t_2} F(t)\ F(t)^T\ dt.$$ Then $\underline{f}_i(t)$ are linearly independent on $[t_1, t_2]$ if and only if the n×n constant

matrix $W(t_1, t_2)$ is nonsingular.

Proof: The proof of this lemma can be found in [Chen84].

Theorem 1 (Uniqueness of particle motion)

If the instantaneous velocity $\underline{V}(t)$ of the moving particle is modeled as truncated Taylor series up to degree n_X, n_Y and n_Z, respectively, and x(t) and y(t) in (2.1.5) are irreducible and are not equal to zero, then the scaled velocity $\underline{V}'(t)$ ($= \dfrac{1}{Z(0)}\ \underline{V}(t)$) can be recovered uniquely from x(t) and y(t).

Proof: From (2.1.5), we have

$$\sum_{i=1}^{n_X+1} V'_{X(i-1)}\ \frac{t^i}{i} - x(t) \sum_{i=1}^{n_Z+1} V'_{Z(i-1)}\ \frac{t^i}{i} = x(t) - x(0) \tag{2.1.11a}$$

$$\sum_{i=1}^{n_Y+1} V'_{Y(i-1)}\ \frac{t^i}{i} - y(t) \sum_{i=1}^{n_Z+1} V'_{Z(i-1)}\ \frac{t^i}{i} = y(t) - y(0) \tag{2.1.11b}$$

In matrix form, (2.1.11) can be rewritten as

$$[V'_{X0}\ \cdots\ V'_{Xn_X}\ V'_{Y0}\ \cdots\ V'_{Yn_Y}\ V'_{Z0}\ \cdots\ V'_{Zn_Z}]\begin{bmatrix} \underline{f}_1(t) \\ \underline{f}_2(t) \\ \cdot \\ \\ \underline{f}_{n_X+n_Y+n_Z+3}(t) \end{bmatrix} = [x(t) - x(0)\ y(t) - y(0)] \tag{2.1.12}$$

where $\underline{f}_i(t)$ i= 1, ... , $n_X + n_Y + n_Z + 3$ are defined in (2.1.6). Multiplying $F^T(t)$, on both sides and integrating from t_1 to t_2, we obtain

$$[V'_{X0}\ \cdots\ V'_{Xn_X}\ V'_{Y0}\ \cdots\ V'_{Yn_Y}\ V'_{Z0}\ \cdots\ V'_{Zn_Z}]\ W(t_1, t_2) = \int_{t_1}^{t_2} [x(t) - x(0)\ y(t) - y(0)]\ F^T(t)\ dt \tag{2.1.13}$$

where $W(t_1, t_2)$ is defined in lemma 2. From Lemma 1 and 2, the matrix $W(t_1, t_2)$ is nonsingular. Consequently, we have

$$[V'_{X0}\ \cdots\ V'_{Xn_X}\ V'_{Y0}\ \cdots\ V'_{Yn_Y}\ V'_{Z0}\ \cdots\ V'_{Zn_Z}] = [\int [x(t)-x(0)\ y(t)-y(0)]\ F^T(t)\ dt]\ W^{-1}(t_1, t_2) \tag{2.1.14}$$

Thus, from (2.1.3), $\dfrac{1}{Z(0)}\underline{V}(t)$ can be recovered uniquely.   Q.E.D.

There are some degenerate cases of x(t) and y(t) for which we can not recover the velocity uniquely. If x(t)=y(t)=0, then $V_X(t) = V_Y(t) = 0$ and $V_Z(t)$ can be arbitrary. In this case, the particle moves along the optical axis. If x(t)=x(0) and y(t)=y(0), then velocities $1/Z(0)\,\underline{V}(t)$ with $V_Z(t) = X(0)\ V_X(t) = Y(0)\ V_Y(t)$ are the solutions which satisfy (2.1.5). In this case, the x(t) and y(t) are reducible. Note that the above two cases are the only degenerated cases for the the motion of constant velocity. For the higher order model of the velocity, we will study further the physical interpretation of the solutions when the rational functions in (2.1.5) are reducible.

2.2 ESTIMATION FROM NOISY MEASUREMENTS

If there is no noise in the measurements of $[x(t)\ y(t)]^T$, then we can use (2.1.14) to find the 3-D instantaneous velocity of the moving particle. However, in practice, we can only get the noisy measurements of the projective position at discrete times. In this section, we address the problem of finding the motion of a moving particle from these measurements. We will discuss the importance of (2.1.5) in estimating the motion and then apply the method of Maximum Likelihood to find the estimate.

As we discussed in the previous section, if we model the instantaneous velocity $\underline{V}(t)$ as a truncated Taylor series, then (2.1.5) gives the explicit relation between the projective images $[x(t)\ y(t)]$ and the unknown parameters: the coefficients in (2.1.3) scaled by $1/Z(0)$, i.e., $\{V'_{Xi}, i = 0, 1, ..., n_X\}$, $\{V'_{Yi}, i = 0, 1, ..., n_Y\}$, and $\{V'_{Zi}, i = 0, 1, ..., n_Z\}$. This means that this relation is regressive [Spre69, Will59]. This observation is useful in two aspects. First, as many regression models do, we can use the previous estimates of the parameters to predict where the next projective position will be. This is important in obtaining the reliable measurements of the projective position because this prediction will reduce the size of the space to be searched for the moving features on the image plane and decrease the chance of mis-matching the feature. Second, we can apply any of the standard approaches in finding the estimates for regression relations and the analysis for the performance of these estimates to our motion analysis. In the following, the method of Maximum Likelihood is discussed to estimate the motion.

Let $[x_n(t_j)\ y_n(t_j)]^T$ be the noisy measurements of $[x(t)\ y(t)]$ at time t_j, j=0,...,J, where (J+1) is the number of avaliable frames. Define the motion parameter vector $\underline{\theta}$ and the measurement vector $\underline{M}$ as follows.

$$\underline{\theta} = [V'_{X0}\ \cdots\ V'_{Xn_X}\ V'_{Y0}\ \cdots\ V'_{Yn_Y}\ V'_{Z0}\ \cdots\ V'_{Zn_Z}]^T \tag{2.2.1}$$

$$\underline{M} = [x_n(t_1)\ y_n(t_1)\ x_n(t_2)\ y_n(t_2)\ \cdots\ x_n(t_J)\ y_n(t_J)]^T \tag{2.2.2}$$

Let $f(\underline{M}|\underline{\theta})$ be the joint probability density function of $\underline{M}$ which are parametrized by $\underline{\theta}$. For some observation of the projective positions, $\underline{M}$ is equal to some fixed value $\underline{m}$ and $f(\underline{M}|\underline{\theta})$ is a function of θ. This function is called the likelihood function and we denote it as $L(\underline{\theta},\underline{m})$, i.e., $L(\underline{\theta},\underline{m}) = f(\underline{M} = \underline{m}|\underline{\theta})$. The method of Maximum Likelihood is to seek $\hat{\underline{\theta}}(\underline{m})$ which satisfies $L(\hat{\underline{\theta}}(\underline{m}),\underline{m}) = \max_{\underline{\theta}}\{L(\underline{\theta},\underline{m})\}$. The interpretation of this estimate is that $\hat{\underline{\theta}}(\underline{m})$ is the parameter most likely to have produced the measurement $\underline{m}$ [Spre69].

Let us model the noise in the measurements $x_n(t_j)$ and $y_n(t_j)$, $j=1,...,J$, as white, zero mean Gaussian noise with constant variance, which are uncorrelated, then finding the Maximum Likelihood estimate of $\underline{\theta}$ is equivalent to solving the least squares problem. That is, if we define the sum of squares function

$$S = \sum_{j} [x_n(t_j) - x(t_j)]^2 + [y_n(t_j) - y(t_j)]^2 \qquad (2.2.3)$$

then the unknown parameters $\{V'_{Xi}, i = 0, 1, ..., n_X\}$, $\{V'_{Yi}, i = 0, 1, ..., n_Y\}$, and $\{V'_{Zi}, i = 0, 1, ..., n_Z\}$ can be found by minimizing S over the space of parameters. Note that the regression relations $x(t)$ and $y(t)$ in (2.1.5) are nonlinear functions with respect to the unknown parameters. Nonlinear optimization can be used in finding the estimates. There are $2 \times (J+1)$ measurements and $n_X + n_Y + n_Z + 3$ unknowns in this optimization. In order to improve the estimate, we may consider the $x(t_0)$ and $y(t_0)$ in (2.1.5) as the additional unknowns. A number of experiments which dealt with the synthetic measurements and real images were conducted and will be discussed in next section.

2.3 EXPERIMENTAL RESULTS

2.3.1. Experiments on synthetic measurements

In these experiments, we use (2.1.5) to generate the exact trajectory of a moving particle by entering the 3-D position at time zero and the velocity of the particle. Adding white, zero mean Gaussian noise to these exact data, we obtain the noisy trajectory as our measurements $x_n(t)$ and $y_n(t)$. We then estimate the velocity by solving the least squares problem in (2.2.3). We use the "Zxssq" routine of IMSL package to obtain the estimate. Zxssq is based on a modification of the Levenberg-Marquardt algorithm which eliminates the need for explicit derivatives. The initial values of the estimates are set to zero. We consider the case of constant velocity. 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. We consider this observed image as 256 × 256 pixels.

Experiment 1: Different noise level

In this experiment, we compare the estimates at different noise levels. The standard deviations of the noise are 0.5, 2.74 and 5 pixels, respectively. The 3-D velocity is (5.8, 6.2, 20) units/second. The time interval between frames is 0.05 second. The 3-D position of the pariticle at time zero is (0, 0, 20) units. Figure 2a shows the exact and noisy trajectories on the image plane. (Only the trajectory with 2.74 pixels error is shown). Figures 2b and 2c are the x and y components of the trajectory. Figures 2d-2f are the relative error in percentage of the estimated V_X', V_Y' and V_Z'. From these results, we have the following observations. The estimated errors decrease as the noise in the measurements decrease. The estimated errors for V_X' and V_Y' are smaller than that for V_Z' as expected. As we use more frames in finding the estimate, the estimated error decreases. Note that if we only use three frames, which is the minimal number of frames for determining the constant motion, to find the estimate, the estimated error is very large. This justifies our earlier contention that the problem of estimating the motion is ill-conditioned when we only use a small number of frames.

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 as we change the number of frames used in the estimation. The 3-D velocity of the particle is (7.5, -8.8, 30) units/second. The time interval between frames is 0.04 second. The standard deviation of the noise is 2.5 pixels. The 3-D position of the pariticle at time zero is (0, 0, 20) units. Figures 3a, 3b and 3c show the exact and a typical noisy trajectories and their x and y components. Figures 3d-3f show the sample mean as well as ± 1 standard deviation of the estimates versus number of frames used in the estimate. It is observed that the sample mean converges to the true values and the sample standard deviations decrease as we use more frames in our estimates. The estimated errors of V_Z' are larger than those of V_X' and V_Y', and the error is significantly large when we use only a few of frames in finding the estimates.

Experiment 3: Higher order motion

In this experiment, the motion is modeled as constant acceleration. The 3-D velocity and the acceleration of the moving particle at time zero are (20, -16, 30) units/second and (-10, 9, -15) units/second^2, respectively. The time interval between frames is 0.04 seconds and the standard deviations of the noise is 2.74 pixels. The 3-D position of the particle at time zero is (-6, -1, 20) units. Figure 4a shows the exact and noisy trajectories on the image plane. Figures 4b and 4c are the x and y components of the trajectory. Figures 4d-4f are the exact and estimated $V_X'(t)$, $V_Y'(t)$ and $V_Z'(t)$, respectively. From these results, we observed that the estimated velocities scaled by 1/Z(0) are covergent to the true values.

2.3.2. Experiments on real images

In these experiments, we generate the image sequences and get the measurements of the projective positions as follows. We put a CCD camera on a robotic arm and move the arm with known velocity respect to a stationary airplane model. This is equivalent to object motion with the same velocity respect to the camera in the reverse direction. We threshold the observed images and calculate the center of mass of the binary images to yield the projective positions (x(t) y(t)) in pixels. These values are transformed to the world coordinate system as follows.

$$x(mm) = (x(pixels) - x_c) \times xscale \times xyratio, \quad y(mm) = (y(pixels) - y_c) \times yscale$$

where (x_c, y_c) is the center of image, (xscale, yscale) is the conversion factor between the physical image plane and the digitized image, and xyratio is the aspect ratio of camera. Then we normalize the above values with respect to the focal length of the camera to get our measurements. Figure 5 shows the set up of these experiments. Note that only one camera is used. In our camera calibration, x_c, y_c, xscale, yscale and xyratio are 253.53 pixels, 257.20 pixels, 0.023 mm/pixel, 0.0134 mm/pixel and 0.7288, respectively. The focal length of the camera is 8.5 mm. The image is 512×512 pixels in size. We consider the case of constant velocity and the sampling rate is 1 frame/second.

Experiment 4 and 5

Figures 6a and 7a show the images of zero frames overlapped with the threshold images of frames 0, 8, 16, 24, 32, 40 and 48 for experiments 4 and 5, respectively. Figures 6b and 7b show the measurements of x and y components, respectively. Figures 6c and 7c are the estimates of V_X', V_Y' and V_Z' and figures 6d and 7d show the ratio of V_Y' / V_X' and V_Z' / V_X' for these two experiments, respectively.

For experiment 4, the 3-D velocity is (6, 4, -5) mm/second. The z component of the particle at time zero is measured as 46.4 cm. In figure 6c, the estimates converge to the value of (-0.0130, -0.00855, 0.0102), which is equal to (6.0320, 3.9672, -4.7328) mm/second. This is quite close to the input velocity of the camera. The estimates of the ratio of V_Y' / V_X' and V_Y' / V_X' in figure 6d are convergent to the true values. For experiment 5, the 3-D velocity is (4, 4, -4) mm/second. The z component of the particle at time zero is measured as 47.2 cm. In figure 7c, the estimates converge to the value of (-0.008636, -0.008469, 0.007703), which is equal to (4.0762, 3.9974, -3.6358) mm/second. This is still close to the input velocity of the camera. Figure 7d shows that the estimates of the ratio of V_Y' / V_X' and V_Y' / V_X' are also convergent to the true values.

3. MOTION ANALYSIS OF A MOVING RIGID BODY

In this section, we are interested in estimating the motion of a moving rigid body from its image sequence. Section 3.1 studies the model of the motion. Section 3.2 proposes to use the state equation approach to find the regression relations between the unknown motion parameters and the projective positions. The method of Maximum Likelihood is used to estimate the parameters. Experimental results and discussions are given in section 3.3.

3.1 MODEL OF THE RIGID BODY MOTION

Suppose that there are (N+1) feature points on the visible surface of the rigid object and we have measured the projective positions 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 rotation center and the rotation of the whole object with respect to this rotation center. We also know that we can only estimate the translation up to a scale factor because of the perspective projection. Thus, our objective is to find this translation of the rotation center up to a scalar factor and the rotation from the observed projective positions.

Let $\underline{R}_i(t) = [X_i(t)\ Y_i(t)\ Z_i(t)]^T$, $[x_i(t)\ y_i(t)]^T$ and $\underline{V}_i(t) = [V_{iX}(t)\ V_{iY}(t)\ V_{iZ}(t)]^T$ be the 3-D position, the projective position and the instantaneous velocity of i-th feature point at time t, respectively. Let $\underline{V}_0'(t)$ be the velocity of zero-point $\underline{V}_0(t)$ scaled by $1/Z_0(0)$. Then we have

$$\frac{d}{dt}\underline{R}_i(t) = \underline{V}_i(t) \tag{3.1.1}$$

$$x_i(t) = X_i(t)\ /\ Z_i(t), \quad y_i(t) = Y_i(t)\ /\ Z_i(t) \tag{3.1.2a,b}$$

Integrating (3.1.1) from 0 to t, we can find the $\underline{R}_i(t)$ in terms of $\underline{V}_i(t)$, i.e.,

$$\underline{R}_i(t) = \underline{R}_i(0) + \int_0^t \underline{V}_i(\tau)\ d\tau \tag{3.1.3}$$

Let $\underline{\Omega}(t) = [\Omega_X(t)\ \Omega_Y(t)\ \Omega_Z(t)]^T$ be the angular velocity of the motion. From the assumption concerning a rigid body, we have

$$\underline{V}_i(t) - \underline{V}_0(t) = \underline{\Omega}(t) \times (\underline{R}_i(t) - \underline{R}_0(t)) \tag{3.1.4}$$

where $\underline{R}_0(t)$ is the rotation center. We can express (3.1.4) alternatively in the following form.

$$\frac{d}{dt}(\underline{R}_i(t) - \underline{R}_0(t)) = A(t)(\underline{R}_i(t) - \underline{R}_0(t)) \tag{3.1.5}$$

$$\text{where} \quad A(t) = \begin{bmatrix} 0 & -\Omega_Z(t) & \Omega_Y(t) \\ \Omega_Z(t) & 0 & -\Omega_X(t) \\ -\Omega_Y(t) & \Omega_X(t) & 0 \end{bmatrix} \tag{3.1.6}$$

Similarly to the motion of a particle, we will model the velocity of a rigid body as the Taylor series.

The key step of our approach in estimating the rigid body motion is to derive the explicit relation between the projective positions and the unknown translation $\underline{V}_0'(t)$ and the rotation $\underline{\Omega}(t)$. The derivation of this relation is given in next section.

3.2 STATE EQUATION APPROACH IN FINDING THE REGRESSION RELATION

In this section, we describe the state equation approach to finding the regression relations between the unknown motion parameters and the projective positions. Basically, we are going to solve the equation of rigid body motion in (3.1.5). We then propose to use the method of Maximum Likelihood to estimate the motion parameters. By using these regression relations, 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. We first consider the case of constant translation and rotation and then extend the study to general motion. We also investigate how to reduce the number of unknowns if all the points lie on a planar surface.

3.2.1 CONSTANT TRANSLATION AND ROTATION

For convenience, we use $\underline{V}(t)$ instead of $\underline{V}_0(t)$ to represent the translation of the rotation center. If the translation and the rotation of a moving rigid body are constant, i.e.,

$$\underline{V}(t) = [V_{X0}\ V_{Y0}\ V_{Z0}]^T, \quad \underline{\Omega}(t) = [\Omega_{x0}\ \Omega_{y0}\ \Omega_{z0}]^T \quad (3.2.1a,b)$$

then the projective position of the rotation center is given, from (3.1.2) and (3.1.3), as follows.

$$x_0(t) = \frac{x_0(0) + V_{X0}'\ t}{1 + V_{Z0}'\ t}, \quad y_0(t) = \frac{y_0(0) + V_{X0}'\ t}{1 + V_{Z0}'\ t} \quad (3.2.2a,b)$$

where $[V_{X0}'\ V_{Y0}'\ V_{Z0}']^T = \frac{1}{Z_0(0)}[V_{X0}\ V_{Y0}\ V_{Z0}]^T$ is the translation velocity of the rotation center with the

scalae factor $1 / Z_0(0)$. Because of the constant rotation, (3.1.5) is a state equation of the state vector $[\underline{R}_i(t) - \underline{R}_0(t)]$ with the constant matrix A. Its solution is

$$\underline{R}_i(t) - \underline{R}_0(t) = e^{A(t-t_0)}(\underline{R}_i(t_0) - \underline{R}_0(t_0)) \quad (3.2.3)$$

where $e^{A(t-t_0)}$ is the state transition matrix and e^{At} is equal to the inverse Laplace transform of $(sI - A)^{-1}$. Let the components of e^{At} be α_{ij}, i,j= 1,2,3, then we have

$$\alpha_{11} = \frac{\Omega_X^2}{\omega^2} + \frac{(\Omega_Y^2 + \Omega_Z^2)}{\omega^2}\cos\omega t, \quad \begin{Bmatrix}\alpha_{12}\\\alpha_{21}\end{Bmatrix} = \frac{\Omega_X\Omega_Y}{\omega^2}(1 - \cos\omega t)\begin{Bmatrix}-\\+\end{Bmatrix}\frac{\Omega_Z}{\omega}\sin\omega t \quad (3.2.4a,d)$$

$$\alpha_{22} = \frac{\Omega_Y^2}{\omega^2} + \frac{(\Omega_X^2 + \Omega_Z^2)}{\omega^2}\cos\omega t, \quad \begin{Bmatrix}\alpha_{13}\\\alpha_{31}\end{Bmatrix} = \frac{\Omega_Z\Omega_X}{\omega^2}(1 - \cos\omega t)\begin{Bmatrix}+\\-\end{Bmatrix}\frac{\Omega_Y}{\omega}\sin\omega t \quad (3.2.4b,e)$$

$$\alpha_{33} = \frac{\Omega_Z^2}{\omega^2} + \frac{(\Omega_X^2 + \Omega_Y^2)}{\omega^2}\cos\omega t, \quad \begin{Bmatrix}\alpha_{23}\\\alpha_{32}\end{Bmatrix} = \frac{\Omega_Y\Omega_Z}{\omega^2}(1 - \cos\omega t)\begin{Bmatrix}-\\+\end{Bmatrix}\frac{\Omega_X}{\omega}\sin\omega t \quad (3.2.4c,f)$$

where $\omega = \pm(\Omega_X^2 + \Omega_Y^2 + \Omega_Z^2)^{1/2}$. There are two solutions for the ω in solving the state equation. The absolute values of them are equal to the amplitude of $\underline{\Omega}$. Note that we have the same α_{ij} in (3.2.4) for two different ω's and our unknowns for the rotation are Ω_X, Ω_Y and Ω_Z, not ω. From (3.1.2) and (3.1.3), we have

$$x_i(t) = \frac{x_0(0) + V'_{X0}\, t + \alpha_{11}(x_i(0)\, Z_{i0} - x_0(0)) + \alpha_{12}(y_i(0)\, Z_{i0} - y_0(0)) + \alpha_{13}(Z_{i0} - 1)}{1 + V'_{Z0}\, t + \alpha_{31}(x_i(0)\, Z_{i0} - x_0(0)) + \alpha_{32}(y_i(0)\, Z_{i0} - y_0(0)) + \alpha_{33}(Z_{i0} - 1)} \qquad (3.2.5a)$$

$$y_i(t) = \frac{y_0(0) + V'_{Y0}\, t + \alpha_{21}(x_i(0)\, Z_{i0} - x_0(0)) + \alpha_{22}(y_i(0)\, Z_{i0} - y_0(0)) + \alpha_{13}(Z_{i0} - 1)}{1 + V'_{Z0}\, t + \alpha_{31}(x_i(0)\, Z_{i0} - x_0(0)) + \alpha_{32}(y_i(0)\, Z_{i0} - y_0(0)) + \alpha_{33}(Z_{i0} - 1)} \qquad (3.2.5b)$$

where $Z_{i0} = Z_i(0)/Z_0(0)$, $i=1,...,N$ is the relative depth between the i-th point and 0-th point at time zero.

Equations (3.2.2) and (3.2.5) give the explicit relations between the projective positions $x_i(t)$ and $y_i(t)$ and the unknown parameters V'_{X0}, V'_{Y0}, V'_{Z0}, Ω_{X0}, Ω_{Y0}, Ω_{Z0} and Z_{i0}, $i= 1, ... , N$. That is, they are the regression relations we want to find. Thus, as we discussed the motion of particle in section 2, we can estimate the unknown parameters by using the method of Maximun Likelihood due to these regression relations. Furthermore, if we model the noise in the measurements $x_{ni}(t_j)$ and $y_{ni}(t_j)$ of the projective positions $x_i(t)$ and $y_i(t)$ at time t_j, $i=0,1,...,N$, $j= 0,2,...,J$, where (J+1) is the number of available frames, are white, zero mean Gaussian noise and they are uncorrelated, then finding the Maximum Likelihood estimate is equivalent to solving the following least squares problem. Define the sum of squares function as

$$S = \sum_{i,\,j} [x_{ni}(t_j) - x_i(t_j)]^2 + [y_{ni}(t_j) - y_i(t_j)]^2 \qquad (3.2.6)$$

Then the parameters, V'_{X0}, V'_{Y0}, V'_{Z0}, Ω_{X0}, Ω_{Y0}, Ω_{Z0} and Z_{i0}, $i= 1, ... , N$, can be estimated by minimizing S over the space of parameters. There are $2 \times J \times (N+1)$ measurements and $(6+N)$ unknowns in this optimization problem, so that we need at least three frames to recover the motion. In order to improve the estimates, we may consider the $x_i(t_0)$ and $y_i(t_0)$, $i=0,1,...,N$, in (3.2.5) as additional unknowns. It is important to note that in estimating the rigid body motion by solving the above least squares problem, we have used, simulaneously, all the available frames and feature points on the rigid body. This means that we utilizes all the information in temporal domain as well as the spatial domain to find our estimates.

3.2.2 MOTION OF PLANAR SURFACE

If all the observed points lie on a planar surface, we can reduce the unknown parameters as follows. At time zero, suppose all the points lay on the planar surface,

$$Z = p\, X + q\, Y + r \qquad (3.2.7)$$

where p and q are the slopes of the planar surface, and r is the distance between the planar surface and the projective center along the optical axis. Then we have

$$Z_i(0) = p \ X_i(0) + q \ Y_i(0) + r \quad i = 0,1,...,N \tag{3.2.8}$$

This implies

$$Z_i(0) = r \ [1 - p \ x_i(0) - q \ y_i(0)]^{-1}$$

$$Z_{i0} = Z_i(0)/Z_0(0) = \frac{1 - px_0(0) - qy_0(0)}{1 - px_i(0) - qy_i(0)} \tag{3.2.9}$$

Thus the unknown parameters Z_{i0}, i=1,...,N, are reduced to p and q. Note that the above derivation is still valid even for the general motion disucussed in the next section.

3.2.3 GENERAL MOTION

If the motion of the rotation center is not constant, we can model its translational velocity by using a truncated Taylor series as we discussed the motion of particles in section 2, i.e.,

$$V_X(t) = \sum_{i=0}^{n_X} V_{Xi} \ t^i, \quad V_Y(t) = \sum_{i=0}^{n_Y} V_{Yi} \ t^i, \quad V_Z(t) = \sum_{i=0}^{n_Z} V_{Zi} \ t^i \tag{3.2.10a,b,c}$$

Then we have

$$x_0(t) = \frac{x_0(0) + \sum_{i=0}^{n_X} V'_{Xi} \ t^{i+1}/(i+1)}{1 + \sum_{i=0}^{n_Z} V'_{Zi} \ t^{i+1}/(i+1)}, \quad y_0(t) = \frac{y_0(0) + \sum_{i=0}^{n_Y} V'_{Yi} \ t^{i+1}/(i+1)}{1 + \sum_{i=0}^{n_Z} V'_{Zi} \ t^{i+1}/(i+1)} \tag{3.2.11a,b}$$

where V'_{Xi}, V'_{Yi} and V'_{Zi} are the coefficients in the expansion with the scale factor $1 / Z_0(0)$. If the rotation $\underline{\Omega}(t)$ is not constant, then the state equation of $R_i(t) - R_0(t)$ in (3.1.5) is time-varying and its solution is as follows.

$$R_i(t) - R_0(t) = \Phi(t, t_0) \ (R_i(t_0) - R_0(t_0)) \tag{3.2.12}$$

where $\Phi(t, t_0)$ is the state transition matrix of the state equation and it is the unique solution of the matrix equation

$$\frac{\partial}{\partial t} \ \Phi(t, t_0) = A(t) \ \Phi(t, t_0) \tag{3.2.13}$$

with the initial condition $\Phi(t_0, t_0) = I$. Unfortunately, there is in general no close-form expression for $\Phi(t, t_0)$ in terms of $A(t)$ except for some special cases [Chen84]. Numerical methods are needed to solve the matrix differential equation (3.2.13) if we want to get the state transition matrix from the matrix $A(t)$. However, we can show that the state transition matrix is nonsingular. From (3.1.6), we know that $A(t)$ is skew symmetric with the components $\Omega_X(t)$, $\Omega_Y(t)$ and $\Omega_Z(t)$. This implies that the state transition matrix is orthonormal. This follows from the following theorem.

Theorem 2

If the time varying matrix $A(t)$ is skew symmetric, i.e., $A^T(t) = -A(t)$, then the state transition matrix $\Phi(t,t_0)$ defined in (3.2.13) is orthonomal, i.e.,

$$\Phi^{-1}(t,t_0) = \Phi^T(t, t_0) \qquad (3.2.14)$$

The proof of this theorem is given in appendix. Consequently, we can find

$$A(t) = [\frac{\partial}{\partial t} \Phi(t, t_0)] \Phi^T(t,t_0) \qquad (3.2.15)$$

Thus, if we can recover the state transition matrix, then we can find the $A(t)$ as well as the rotation $\underline{\Omega}(t)$ from (3.2.15) and (3.1.6). Defining the ij-th components of $\Phi(t, 0)$ as α_{ij}, i,j = 1, 2, 3, we have

$$x_i(t) = \frac{x_0(0) + \sum\limits_{i=0}^{n_X} V_{Xi}' \frac{t^{i+1}}{i+1} + \alpha_{11}(x_i(0) Z_{i0} - x_0(0)) + \alpha_{12}(y_i(0) Z_{i0} - y_0(0)) + \alpha_{13}(Z_{i0} - 1)}{1 + \sum\limits_{i=0}^{n_Z} V_{Zi}' \frac{t^{i+1}}{i+1} + \alpha_{31}(x_i(0) Z_{i0} - x_0(0)) + \alpha_{32}(y_i(0) Z_{i0} - y_0(0)) + \alpha_{33}(Z_{i0} - 1)} \qquad (3.2.16a)$$

$$y_i(t) = \frac{y_0(0) + \sum\limits_{i=0}^{n_Y} V_{Yi}' \frac{t^{i+1}}{i+1} + \alpha_{21}(x_i(0) Z_{i0} - x_0(0)) + \alpha_{22}(y_i(0) Z_{i0} - y_0(0)) + \alpha_{13}(Z_{i0} - 1)}{1 + \sum\limits_{i=0}^{n_Z} V_{Zi}' \frac{t^{i+1}}{i+1} + \alpha_{31}(x_i(0) Z_{i0} - x_0(0)) + \alpha_{32}(y_i(0) Z_{i0} - y_0(0)) + \alpha_{33}(Z_{i0} - 1)} \qquad (3.2.16b)$$

Equations (3.2.11) and (3.2.16) give the explicit relations between the projective positions and the unknown motion parameters α's and V's. Thus, we still find the regression relation for the general motion and we can use the method of Maximum Likelihood as well as the method of least squares to estimate the motion.

3.3 EXPERIMENTAL RESULTS

Similarly to the experiments for the motion of particle, we use (3.2.4) and (3.2.5) to generate the exact trajectory of the moving feature points by entering the 3-D position of the points at time zero, the translation and rotation of the motion. Adding white zero mean Gaussian noise to these exact data, we obtain the noisy trajectories as our measurement. We then estimate the velocity by solving the problem of least squares in (3.2.6). We again use the "Zxssq" routine to obtain the estimate. We utilize all the measurements of different feature points and frames simultaneously. We consider the case of constant velocity. Camera parameters used in the simulation are identical to those used in section 2.3.1.

Experiment 6: Different number of feature points

In this experiment, we compare the estimates when the number of feature points varies. The 3-D positions of the seven points used in the experiment are (0,0,20), (4,-4,20), (-2,-4,20), (4,-3,20), (1.3,1.4,25), (-1.5,2.3,15) and (1,-1,30) units. We use the first three, five, and seven points of them to find the rigid body motion. Zero-th point is the rotation center of the motion. The standard deviations of the noise are 2.5 pixels. The 3-D translation is (-4,0,10) units/second and the angular velocity is (-1,-1,2.5) radian/second. The time interval between frames is 0.04 second. Figure 8a shows the exact and noisy trajectories of the feature points on the image plane. Figures 8b and 8c are the x and y

components of the trajectories, respectively. Figures 8d-8o are the relative error in percentage or estimated error in units/second of the estimated V'_X, V'_Y, V'_Z, Ω_X, Ω_Y, Ω_Z, Z_{10}, Z_{20}, Z_{30}, Z_{40}, Z_{50} and Z_{60}, respectively. From these results, we have the following observation. The estimation errors decrease if we use more points in the estimation process. The estimation errors for the relative depth are very small. As we use more frames in estimating the motion, the estimated error decreases. Still, if we use only a few frames to find the estimate, the error of the estimate is very large even if we use seven points. This also justifies our earlier contention that the problem of estimating the motion is ill-conditioned if we use only a small number of frames.

Experiment 7: 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 as we vary the number of frames. The 3-D positions of the three feature points used in this experiment are (0,0,20), (4,-4,20) and (-2,-4,20) units. The 3-D translation is (-4,3,10) units/second and the angular velocity is (-1.2,1.3,2.3) radians/second. The time interval between frames is 0.04 second. Figure 9a shows the exact and noisy trajectories of the feature points on the image plane. Figures 9b and 9c are the x and y components of the trajectories, respectively. Figures 9d-9k show the sample mean as well as ± one standard deviation of the estimated V'_X, V'_Y, V'_Z, Ω_X, Ω_Y, Ω_Z, Z_{10} and Z_{20} versus number of frames used in the estimation, respectively. From these results, we observe that the sample mean converges to the true values and the sample standard deviations decrease as we use more frames. The error is large when we use only a few frames in finding the estimates.

4. CONCLUDING REMARK

The method we presented in this paper requires relatively large number of frames and the smoothness of 3-D motion during that time span. Since some simple feature points can be extracted almost in real-time using special-purpose hardware, it is not unrealistic to expect to collect the necessary measurements in 1-2 seconds. Also the 3-D velocity in the real world does not change rapidly unless the object is exerted by high-frequency forces. In order to find the estimate faster and to capture the variation of the 3-D motion, we are working on the sequential algorithms in solving our least square problems.

Segmenting feature points into indepently moving objects is an open, but a difficult problem. We are currently working on this problem still following the *temporally-oriented approach*. The 3-D total velocity at individual feature points shall provide better segmentation (as opposed to the segmentation schemes based on the 2-D velocity).

Appendix: Proof of Theorem 2

Taking the transpose on both sides of (3.2.13), we have $\dot{\Phi}^T(t, t_0) = \Phi^T(t, t_0) A^T(t)$. Because A(t) is skew symmetric, $\dot{\Phi}^T(t, t_0) = -\Phi^T(t, t_0) A(t)$. It implies that $\frac{d}{dt}(\Phi^T(t, t_0) \Phi(t, t_0)) = 0$. So, $\Phi^T(t, t_0) \Phi(t, t_0)$ is a constant matrix for all t. But we know that $\Phi(t_0, t_0) = I$, so we have $\Phi(t, t_0)^T \Phi(t, t_0) = I$.   Q.E.D.

Reference

[Agga81]
J.K. Aggarwal, L.S. Davis and W.N. Martin, "Correspondence process in dynamic scene analysis", Proc. IEEE 69, 562-572, 1981.

[Agga85]
J.K. Aggarwal and A. Mitiche, "Structure and motion from images: fact and fiction", IEEE workshop on Computer Vision, Oct 1985, Michigan, pp. 127-128.

[Agga86]
J.K. Aggarwal, "Motion and time-varying imagery-an overview", Proc. Of IEEE Workshop on Motion: representation and analysis, pp 1-6, 1986.

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

[Aloi86]
J. Aloimonos and I. Rigoutsos, "Determining the 3-D motion of a rigid surface patch without correspondce, under perspective projection: I. planar surface II. curved surfaces", AAAI 1986, pp. 681-688.

[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, Jan 1986 , pp 90-99.

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

[Chen84]
C.T. Chen, *Linear system theory and design*, CBS College Publishing, 1984.

[Enke84]
W. Enkelmann, "Investigations of mutigrid algorithms from the estimation of optical flow fields in image sequences", in Proc. IEEE Workshop on motion: representation and analysis, pp.81-87, May 1986.

[Fang84]
J.-Q. Fang and T.S. Huang, "Some experiments on estimating the 3-D parameters of a rigid body from two consecutive image frames", IEEE PAMI, Sept. 1984, pp.545-554.

[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.

[Hara81]
R.M. Haralick and J.S. Lee, "The facet approach to optic flow", in Proceeding of Image Understanding Workshop, (Arlington, VA), pp. 84-93, 1983.

[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, 185-204, 1981.

[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.

[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, pp306-309.

[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.

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

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

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

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

[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, pp554-562.

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

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

[Spre69]
P. Sprent, Models in regression and related topics, Methuen Co. London, 1969.

[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, pp129-140.

[Tsai81]
R.Y. Tsai and T.S. Huang, "Estimating Three-dimensional motion parameters of a rigid planar patch I", IEEE ASSP, Dec 1981, pp 1147-1152.

[Tsai82]
R.Y. Tsai, T.S. Huang and W.L.Zhu, "Estimating three-dimensional motion parameters of a rigid planar patch II : singular value decomposition ", IEEE ASSP Aug 1982, pp 525-534.

[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.

[Tret84]
O. Tretiak and L Pastor, "Velicity estimation from image sequences with second order differential operators", Proc. Int. Conf. of Pattern Recognition, pp. 16-19, 1984.

[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, 72-94, 1985.

[Waxm86a]
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.

[Waxm86b]
A.M. Waxman and K. Wohn, "Contour evolution, neighborhood, deformation and image flow: textured surfaces in motion", in Image understanding '85-86, Richards and Ullman Eds., 1986.

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

[Will59]

E.J. Williams, *Regression Analysis*, John Wiley & Sons, Inc., 1959.

[Wohn83]

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

[Wohn85]

K. Wohn and A.M. Waxman, "Contour evolution, neighborhood deformation and local image flow: curved surfaces in motion", CS TR-1531, Univ. Maryland, July 1985.

[Wu86]

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

[Zhua86]

X. Zhuang, T.S. Huang, and R. M. Haralick, "Two-view motin analysis: A unified algorithm", J. Opt. Sco. Amer., vol. 3, pp. 1492-1500, Sept. 1986.
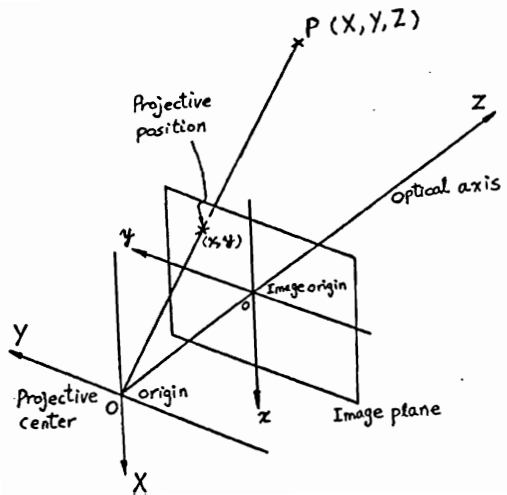
Fig. 1. Camera model and coordinate system.

Exact and noisy trajectories (2.74 pixels error)
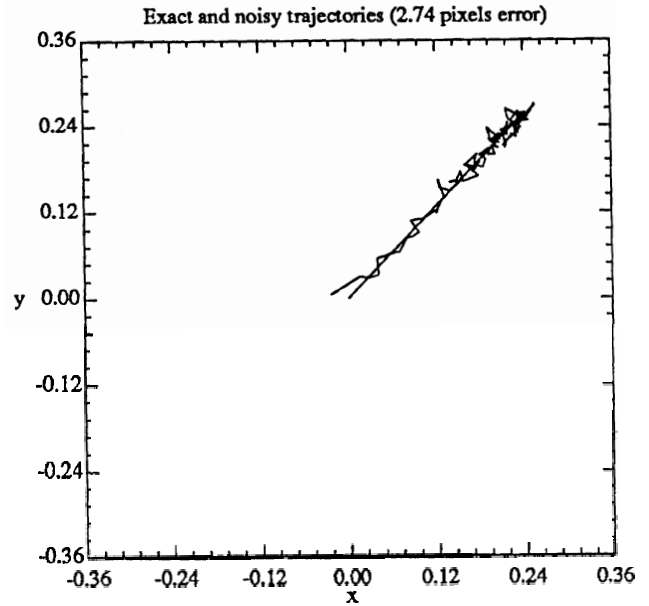
Fig. 2a. Exact and noisy trajectories of the moving particle for experiment 1. Standard deviation of the noise is 2.74 pixels.
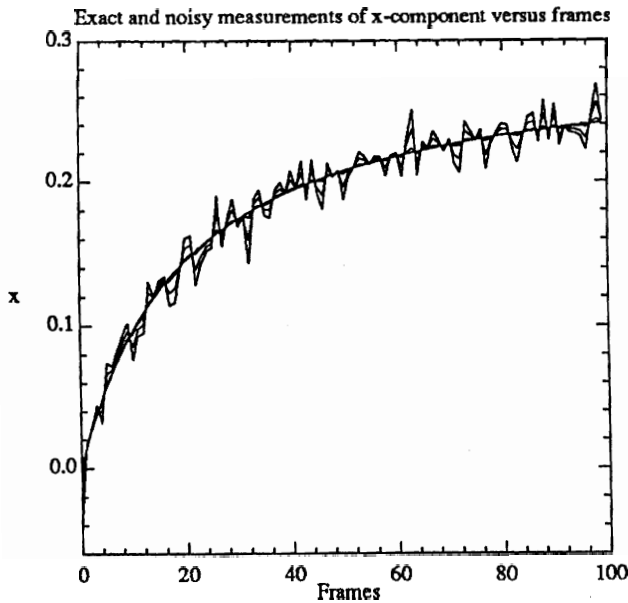
Exact and noisy measurements of x-component versus frames

Fig. 2b. Noisy measurements of x-component of the moving particle versus number of frames for experiment 1. Standard deviations of the noise are 0.5, 2.74 and 5.0 pixels.

Exact and noisy measurements of y-component versus frames
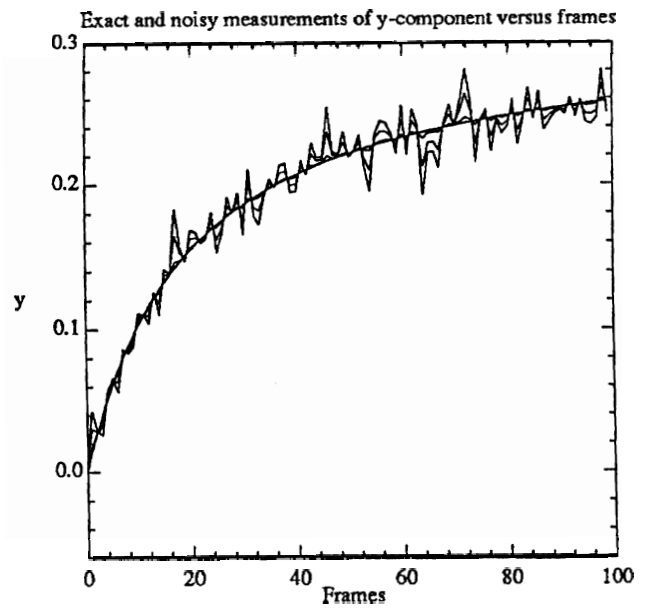
Fig. 2c. Noisy measurements of y-component of the moving particle versus number of frames for experiment 1. Standard deviations of the noise are 0.5, 2.74 and 5.0 pixels.
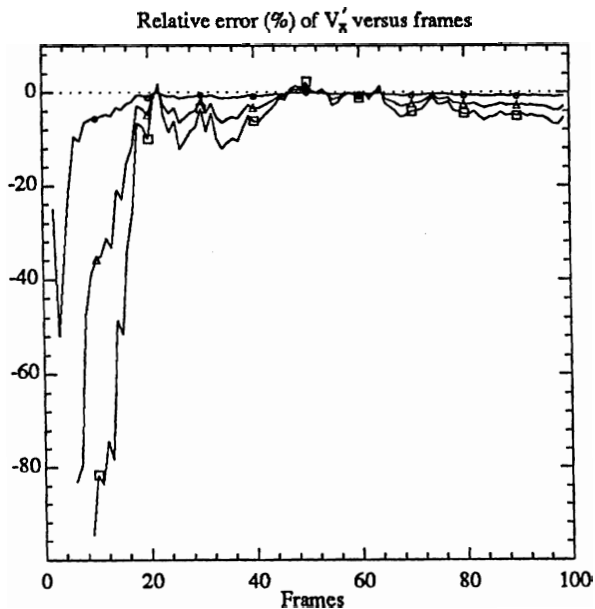
Fig. 2d. Relative error (%) of V_X' versus number of frames for experiment 1. Standard deviations of the noise are 0.5 (O), 2.74(Δ) and 5.0 (□) pixels.
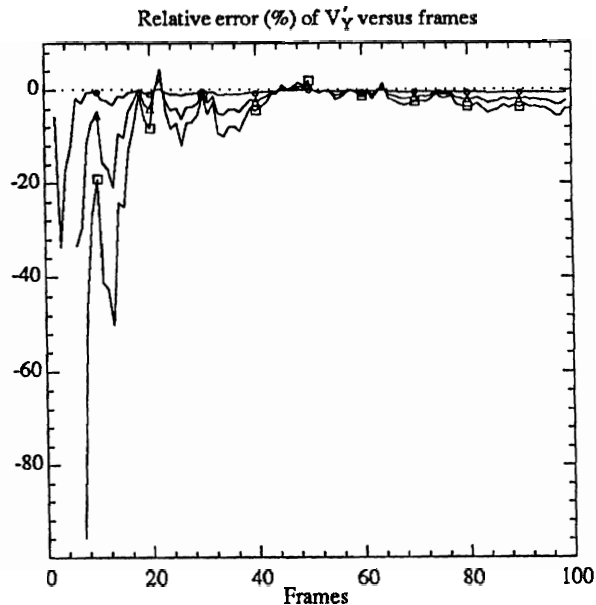
Fig. 2e. Relative error (%) of V_Y' versus number of frames for experiment 1. Standard deviations of the noise are 0.5 (O), 2.74(Δ) and 5.0 (□) pixels.
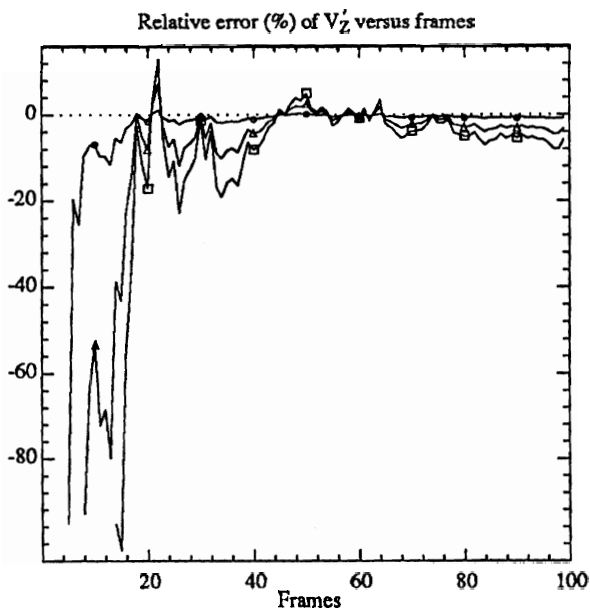
Fig. 2f. Relative error (%) of V_Z' versus number of frames for experiment 1. Standard deviations of the noise are 0.5 (O), 2.74(Δ) and 5.0 (□) pixels.
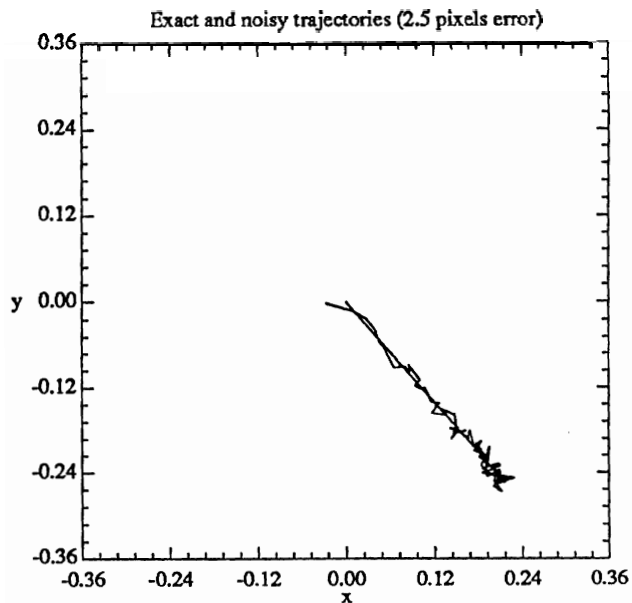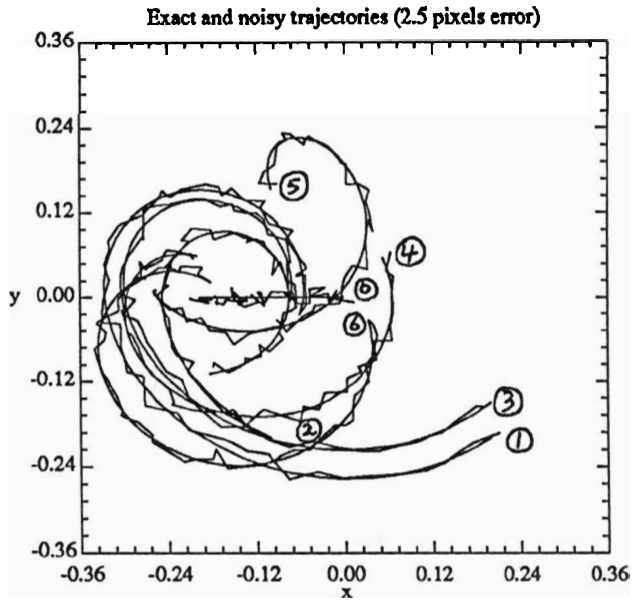

Fig. 3a. Exact and noisy trajectories of the moving particle for experiment 2. Standard deviation of the noise is 2.5 pixels.
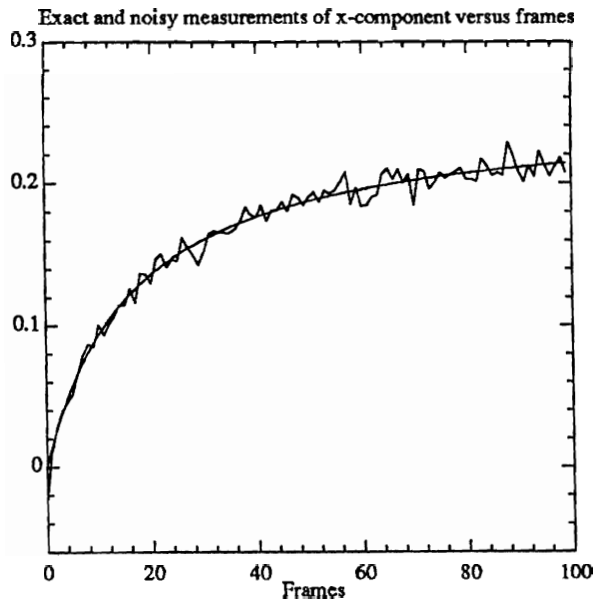
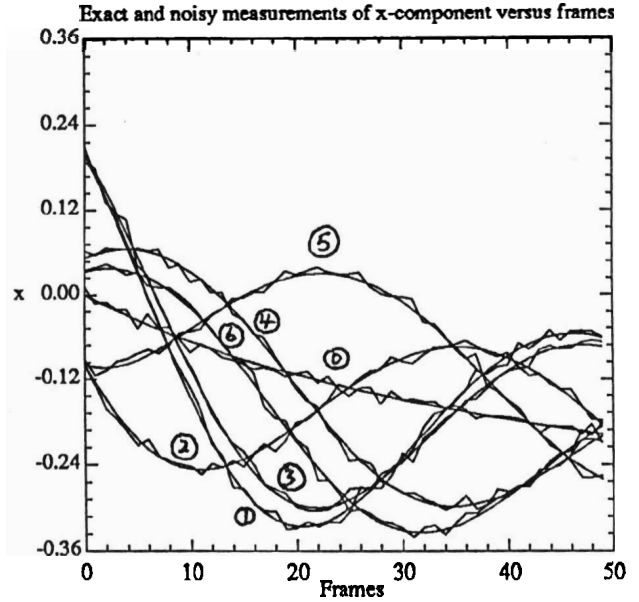Exact and noisy measurements of x-component versus frames

Fig. 3b. Noisy measurements of x-component of the moving particle versus number of frames for experiment 2. Standard deviation of the noise is 2.5 pixels.

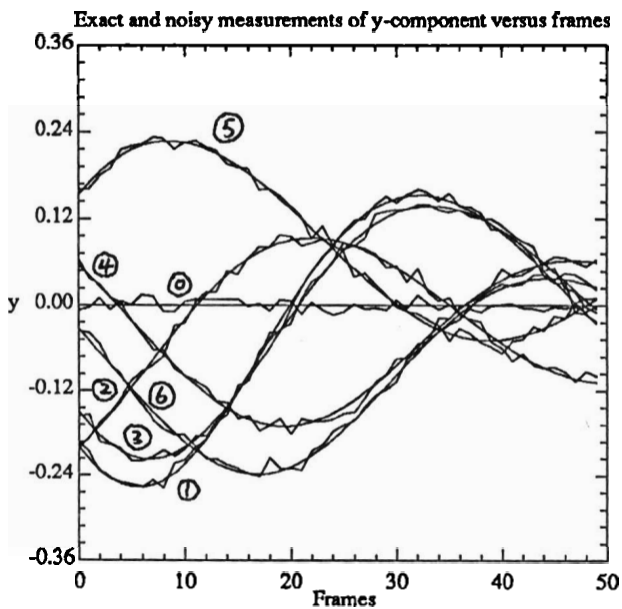Exact and noisy measurements of y-component versus frames

Fig. 3c. Noisy measurements of y-component of the moving particle versus number of frames for experiment 2. Standard deviation of the noise is 2.5 pixels.
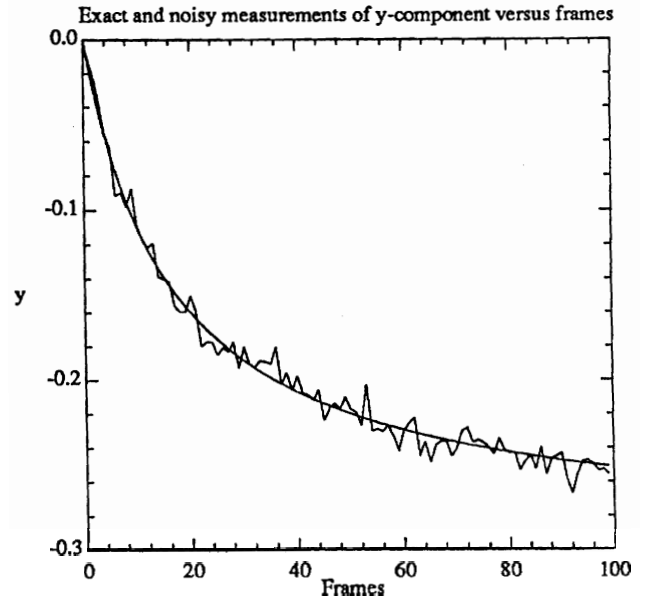
Sample mean and $\pm\sigma$ of V'_X versus frames

Fig. 3d. Sample mean and $\pm$ standard deviation of estimated V'_X versus number of frames for experiment 2.
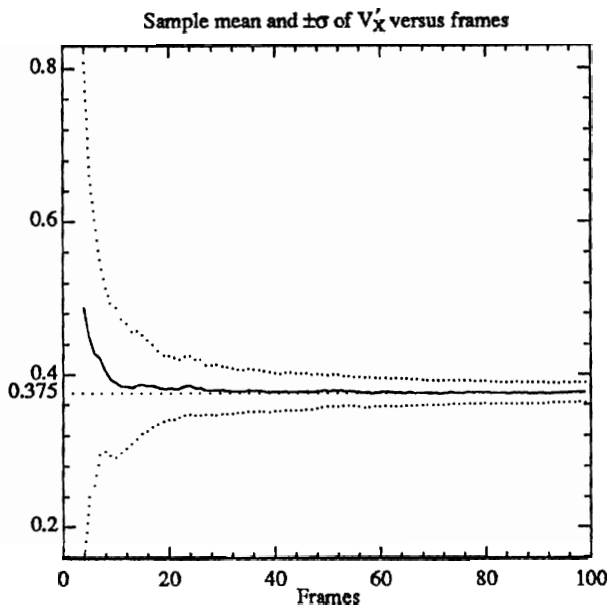
Sample mean and $\pm\sigma$ of V'_Y versus frames .
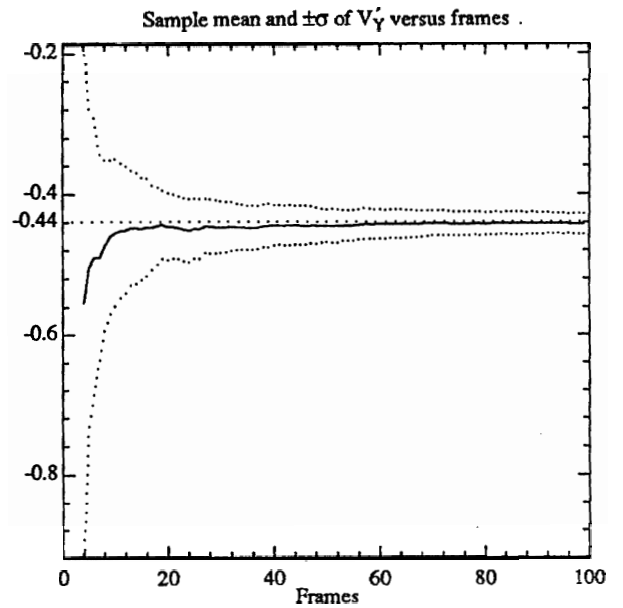
Fig. 3e. Sample mean and $\pm$ standard deviation of estimated V'_Y versus number of frames for experiment 2.
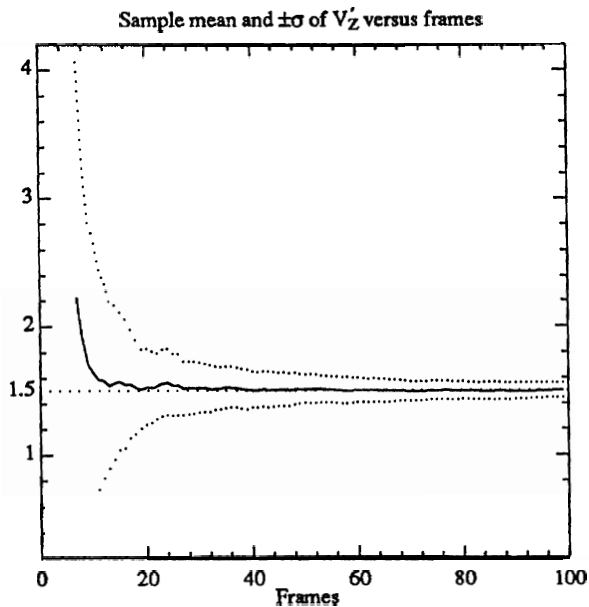
Fig. 3f. Sample mean and ± standard deviation of estimated V_Z' versus number of frames for experiment 2.
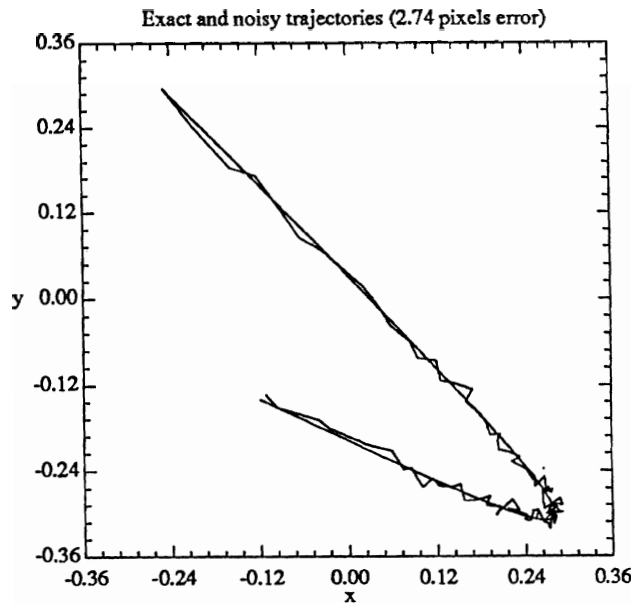
Fig. 4a. Exact and noisy trajectories of the moving particle for experiment 3. Standard deviation of the noise is 2.74 pixels.
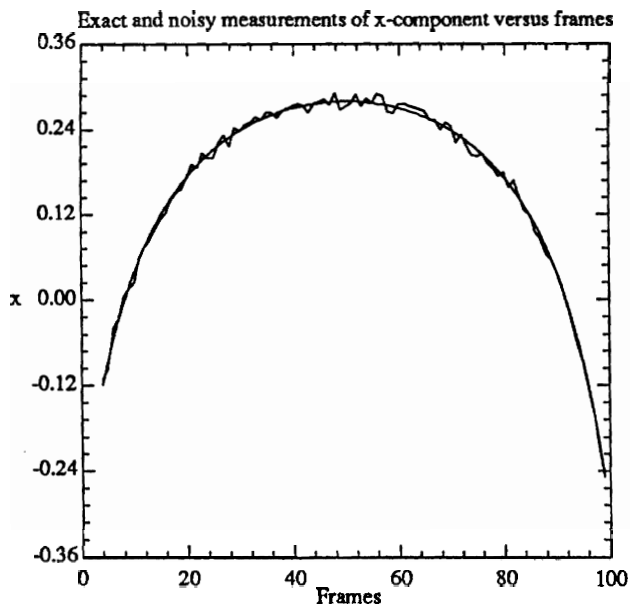
Fig. 4b. Noisy measurements of x-component of the moving particle versus number of frames for experiment 3. Standard deviations of the noise are 2.74 pixels.

Fig. 4c. Noisy measurements of y-component of the moving particle versus number of frames for experiment 3. Standard deviations of the noise are 2.74 pixels.
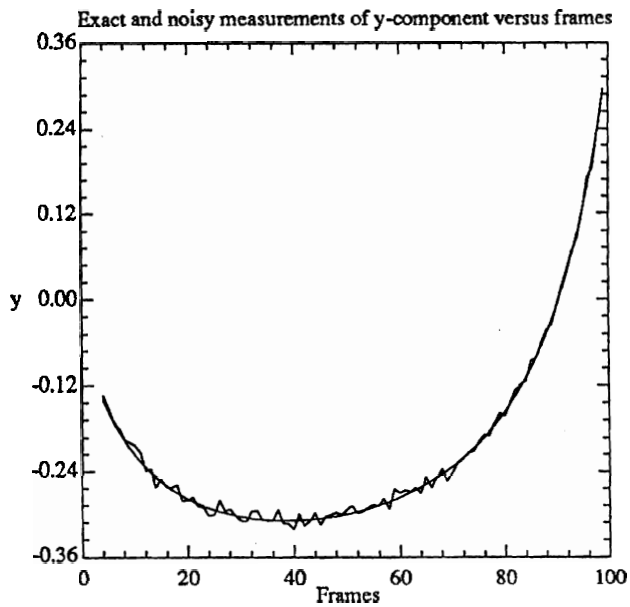
Exact and estimated $V'_X(t)$ versus frames
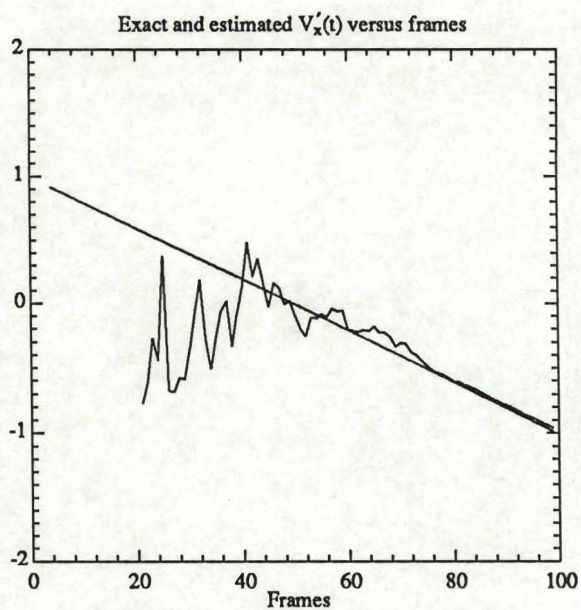
Fig. 4d. Exact and estimated $V'_X(t)$ versus number of frames for experiment 3.

Exact and estimated $V'_Y(t)$ versus frames
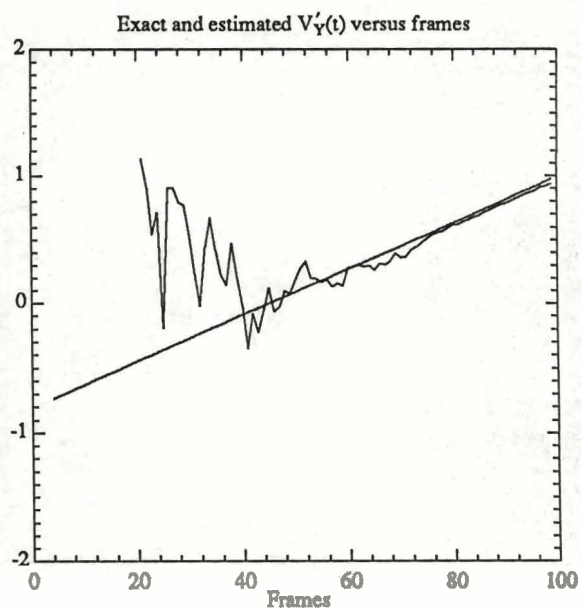
Fig. 4e. Exact and estimated $V'_Y(t)$ versus number of frames for experiment 3.

Exact and estimated $V'_Z(t)$ versus frames
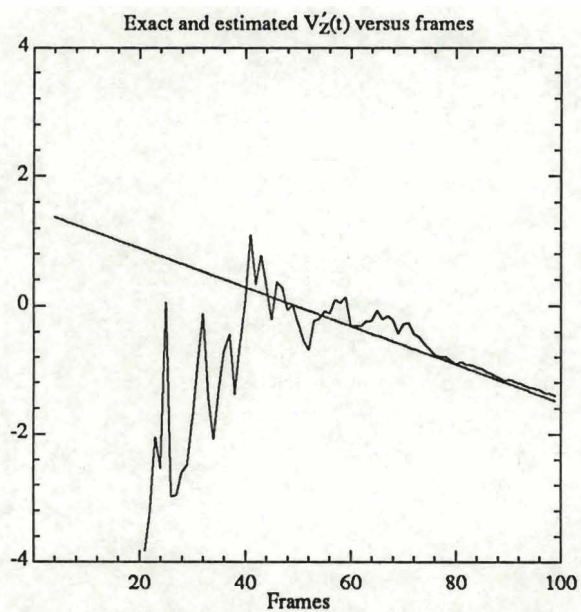
Fig. 4f. Exact and estimated $V'_Z(t)$ versus number of frames for experiment 3.
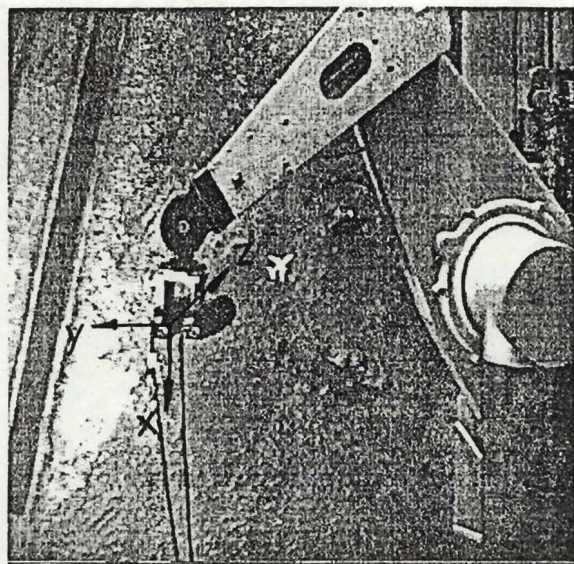
Fig. 5. Setup for real image experiments.
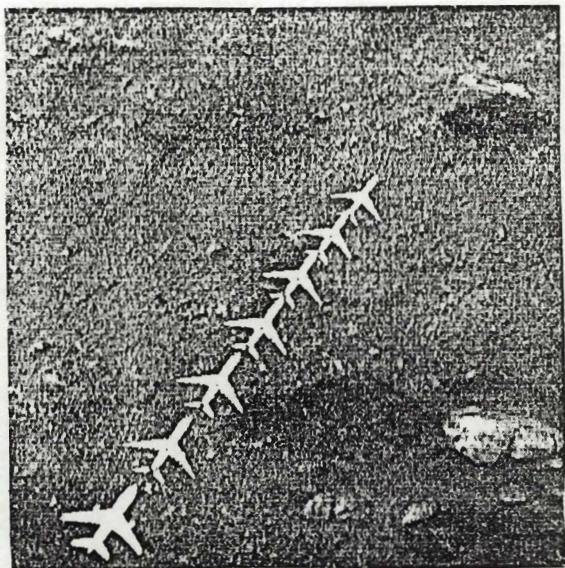
Fig. 6a. Image of zero frame added with threshold images of frames 0, 8, 16, 24, 32, 40 and 48 for experiment 4.
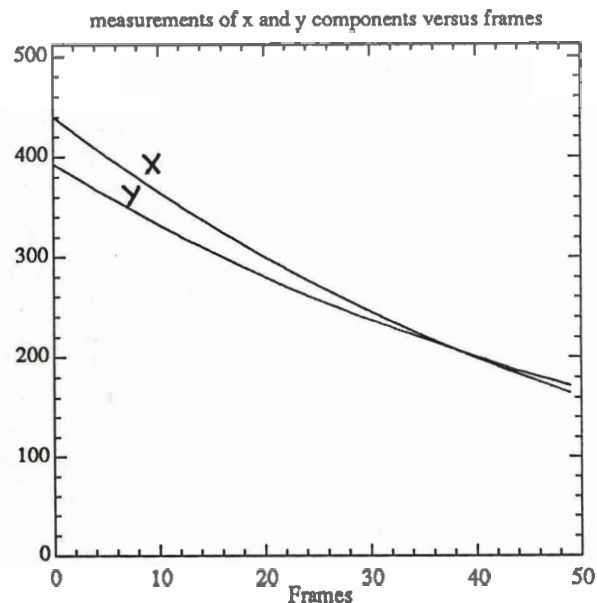
Fig. 6b. Noisy measurements of x and y components (in pixels) versus number of frames for experiment 4.
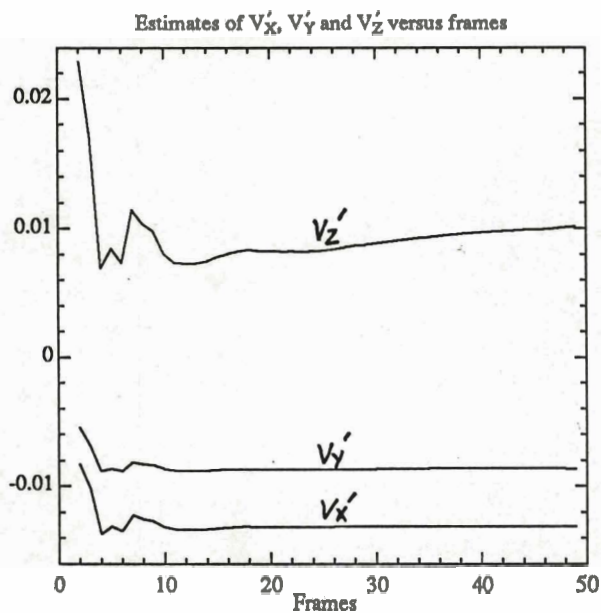
Fig. 6c. Estimates of V_X', V_Y' and V_Z' (in units/second) versus number of frames for experiment 4.
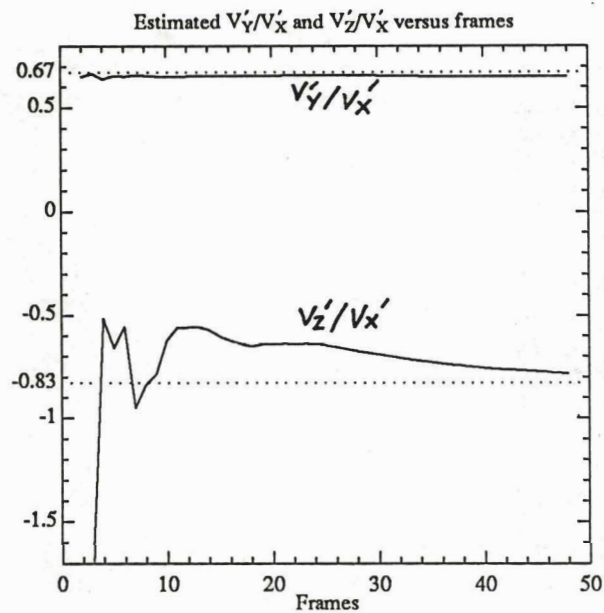
Fig. 6d. Estimated V_Y' / V_X' and V_Z' / V_X' versus number of frames for experiment 4.
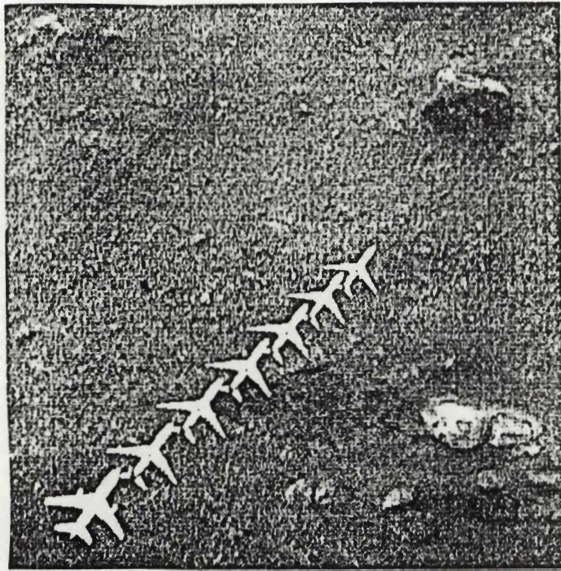
Fig. 7a. Image of zero frame added with threshold images of frames 0, 8, 16, 24, 32, 40 and 48 for experiment 5.
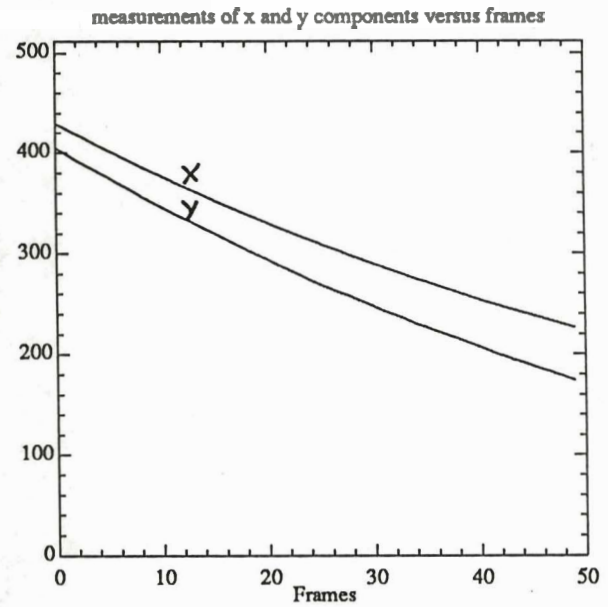
Fig. 7b. Noisy measurements of x and y components (in pixels) versus number of frames for experiment 5.
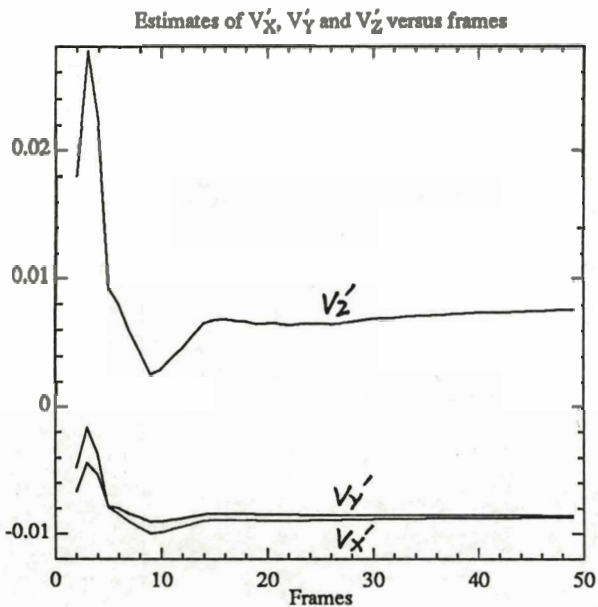
Fig. 7c. Estimates of V'_X, V'_Y and V'_Z (in units/second) versus number of frames for experiment 5.
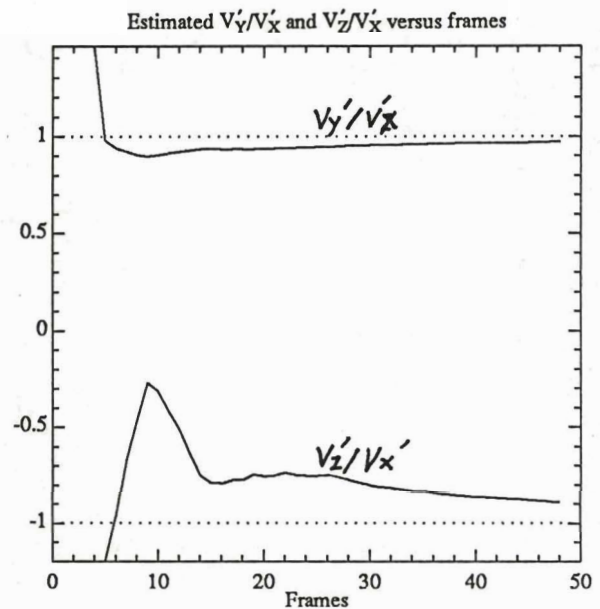
Fig. 7d. Estimated V'_Y / V'_X and V'_Z / V'_X versus number of frames for experiment 5.

Fig. 8a. Exact and noisy trajectories of the moving particles for experiment 6. Standard deviation of the noise is 2.5 pixels.

Fig. 8b. Noisy measurements of x-component of the moving particles versus number of frames for experiment 6. Standard deviation of the noise is 2.5 pixels.

Fig. 8c. Noisy measurements of y-component of the moving particle versus number of frames for experiment 6. Standard deviation of the noise is 2.5 pixels.
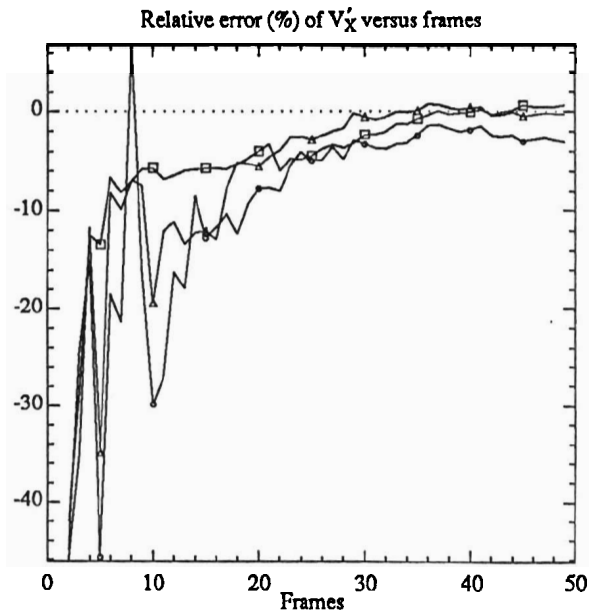
Fig. 8d. Relative error (%) of V_x' versus number of frames for experiment 6 by using three points (O), five points(Δ) and seven points (□).
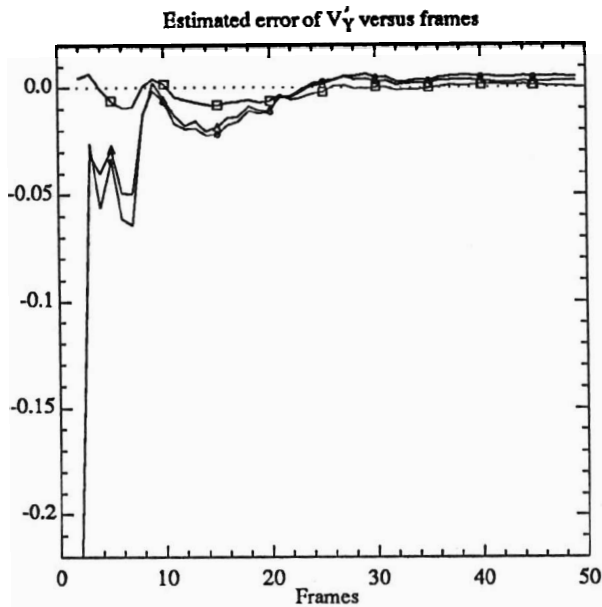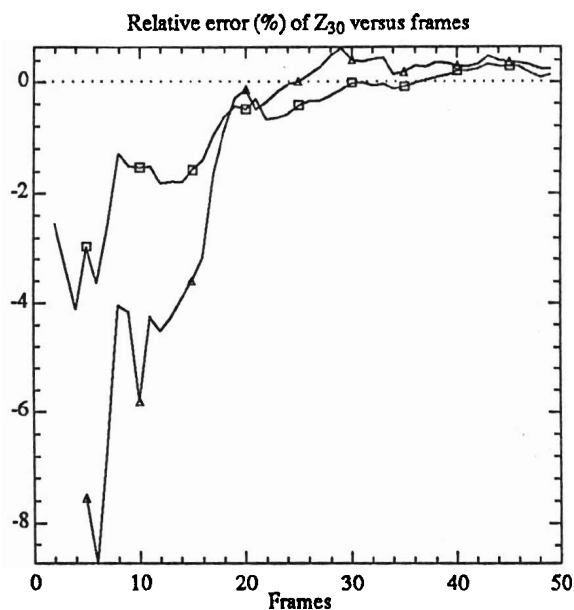
Fig. 8e. Estimated error of V'_Y versus number of frames for experiment 6 by using three points (O), five points(Δ) and seven points (□).
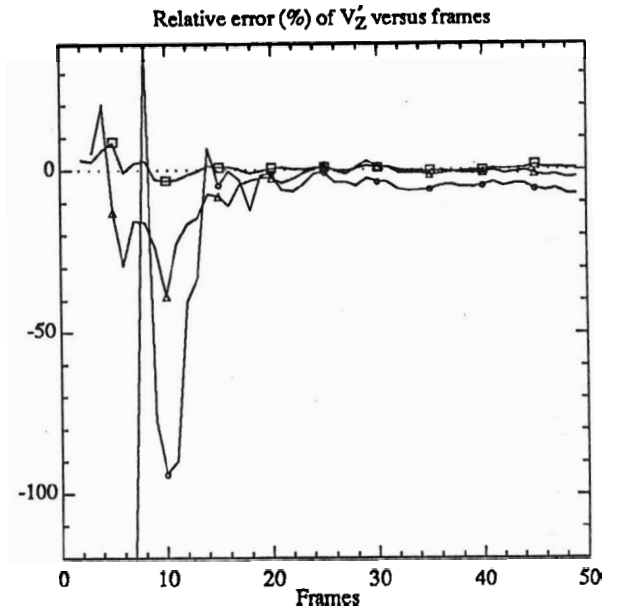
Fig. 8f. Relative error (%) of V'_Z versus number of frames for experiment 6 by using three points (O), five points(Δ) and seven points (□).
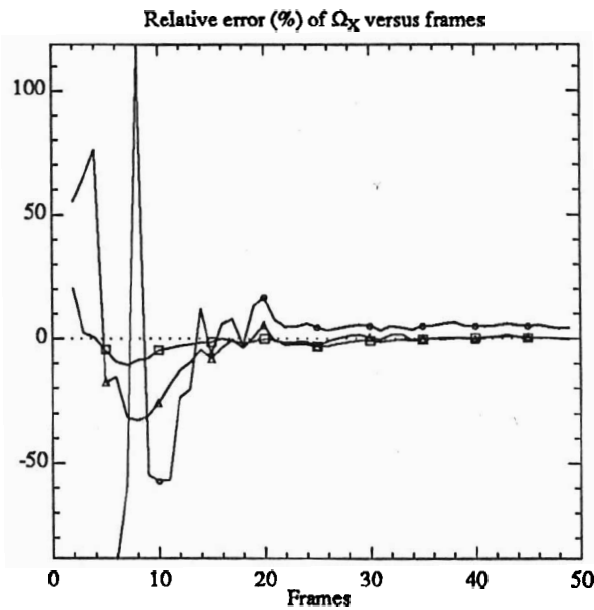
Fig. 8g. Relative error (%) of Ω_X versus number of frames for experiment 6 by using three points (O), five points(Δ) and seven points (□).

Fig. 8h. Relative error (%) of Ω_Y versus number of frames for experiment 6 by using three points (O), five points(Δ) and seven points (□).
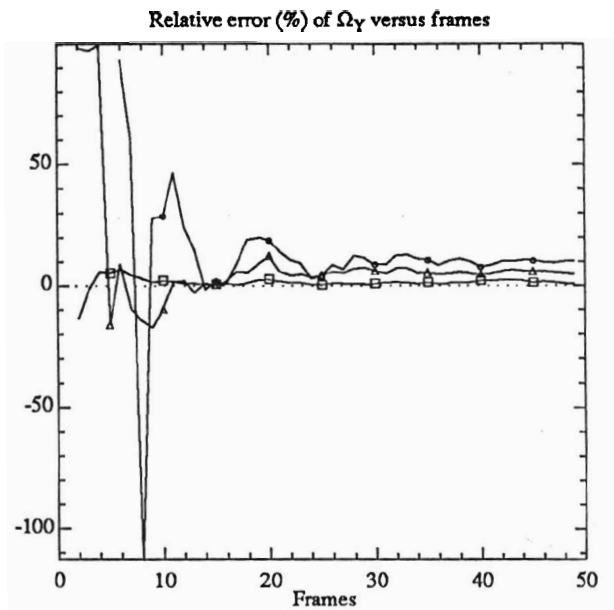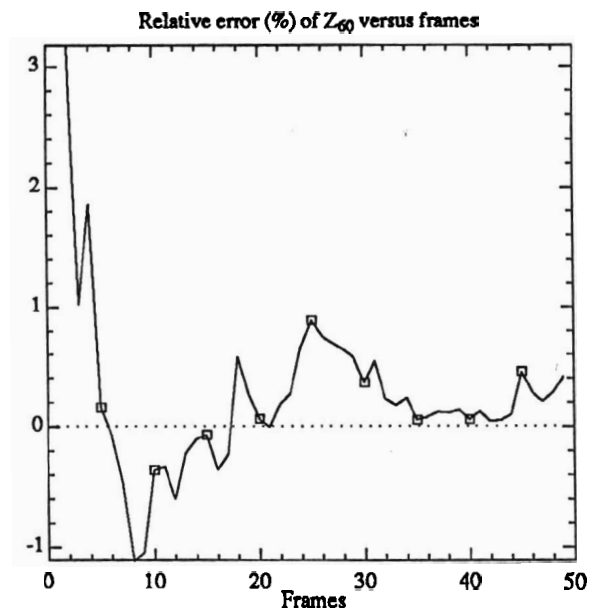
Relative error (%) of Ω_Z versus frames

Fig. 8i. Relative error (%) of Ω_Z versus number of frames for experiment 6 by using three points (O), five points(Δ) and seven points ($\square$).
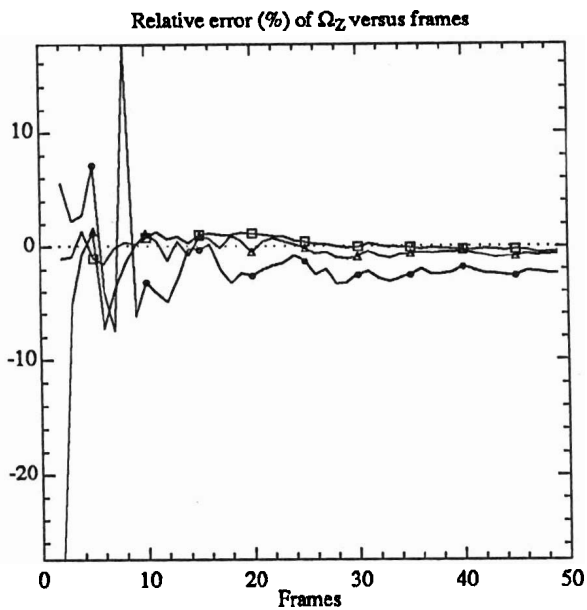
Relative error (%) of Z_{10} versus frames
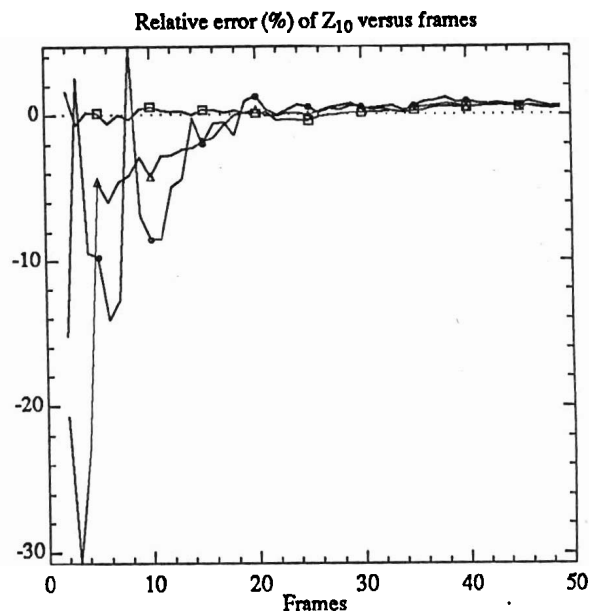
Fig. 8j. Relative error (%) of Z_{10} versus number of frames for experiment 6 by using three points (O), five points(Δ) and seven points ($\square$).

Relative error (%) of Z_{20} versus frames
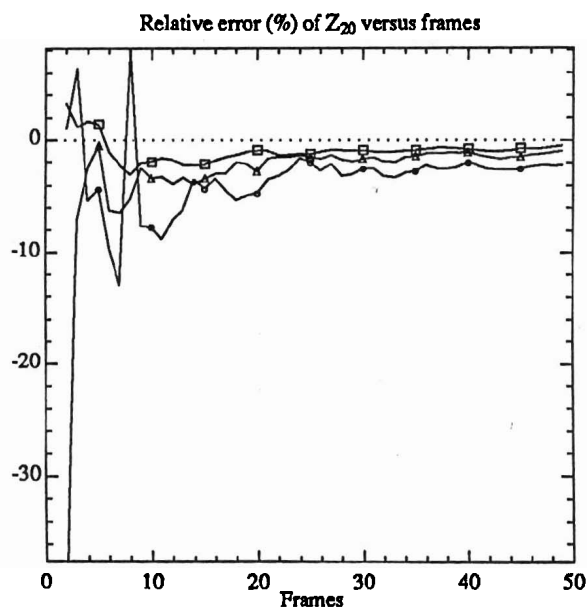
Fig. 8k. Relative error (%) of Z_{20} versus number of frames for experiment 6 by using three points (O), five points(Δ) and seven points ($\square$).

Relative error (%) of Z_{30} versus frames

Fig. 8l. Relative error (%) of Z_{30} versus number of frames for experiment 6 by using five points(Δ) and seven points ($\square$).

Relative error (%) of Z_{40} versus frames
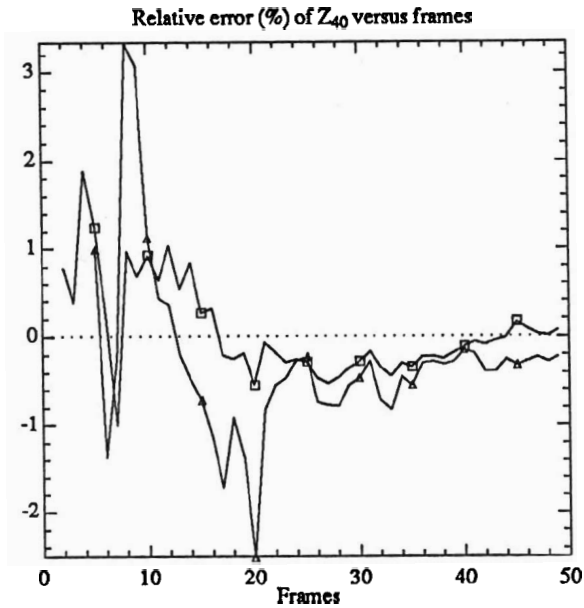
Fig. 8m. Relative error (%) of Z_{40} versus number of frames for experiment 6 by using five points(Δ) and seven points ($\square$).

Relative error (%) of Z_{50} versus frames
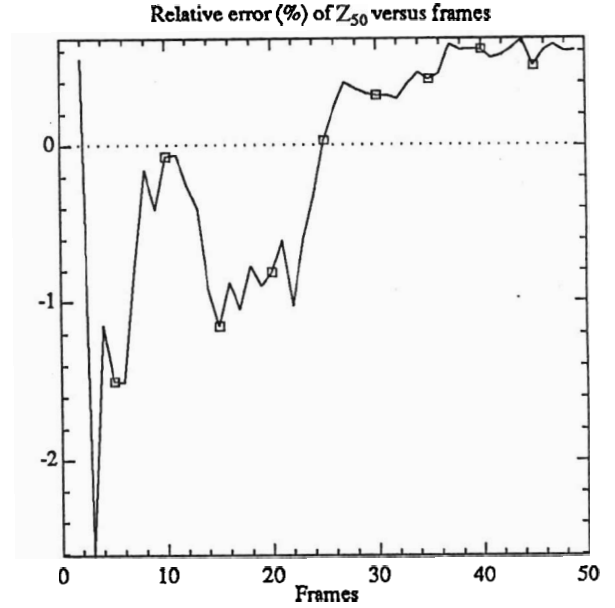
Fig. 8n. Relative error (%) of Z_{50} versus number of frames for experiment 6 by using seven points ($\square$).

Relative error (%) of Z_{60} versus frames

Fig. 8o. Relative error (%) of Z_{60} versus number of frames for experiment 6 by using seven points ($\square$).
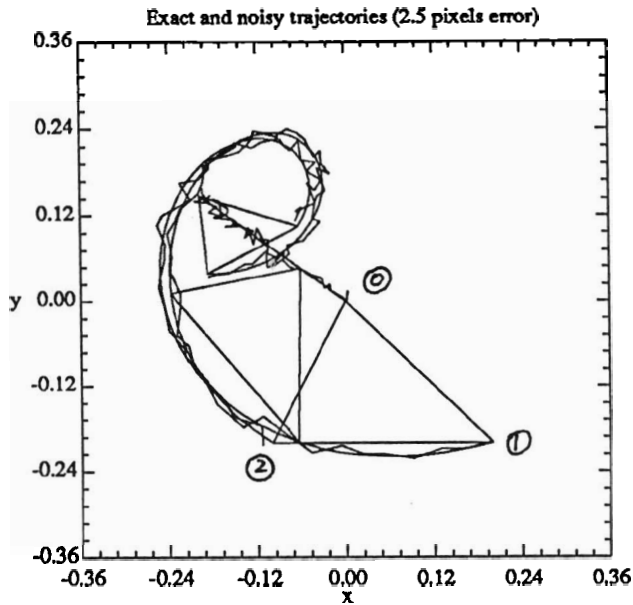
Fig. 9a. Exact and noisy trajectories of the moving particles for experiment 7. Standard deviation of the noise is 2.5 pixels.
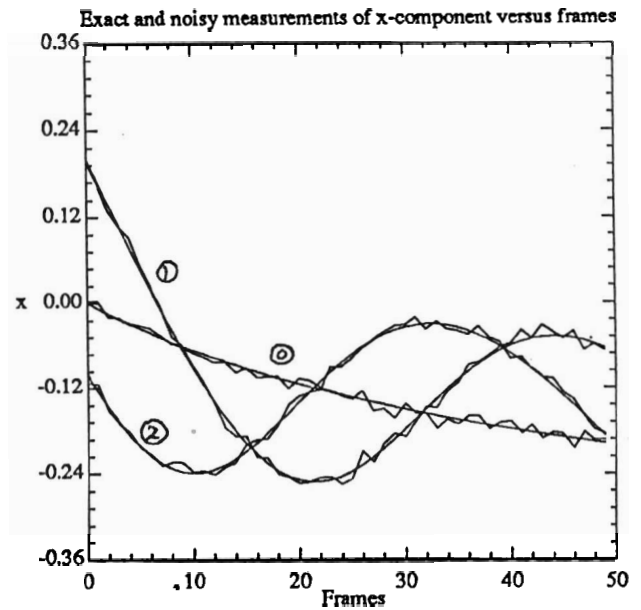
Fig. 9b. Noisy measurements of x-component of the moving particles versus number of frames for experiment 7. Standard deviation of the noise is 2.5 pixels.
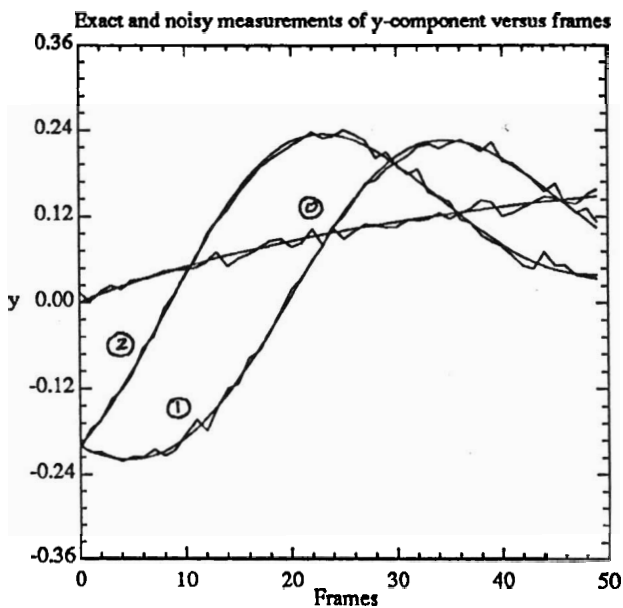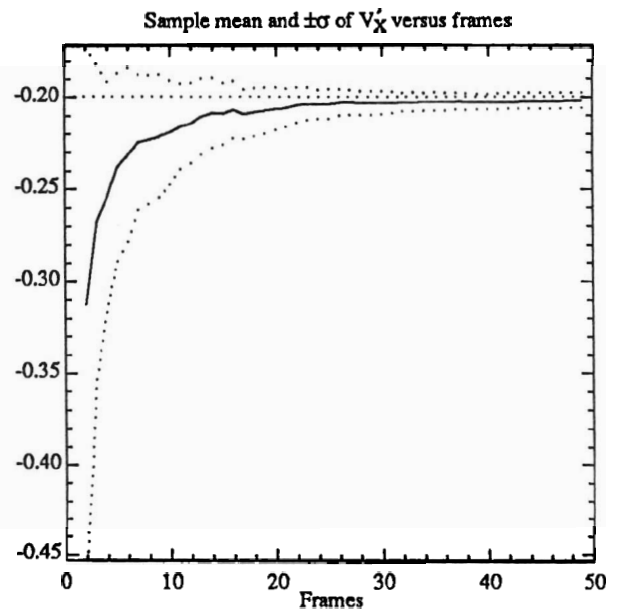

Fig. 9c. Noisy measurements of y-component of the moving particles versus number of frames for experiment 7. Standard deviation of the noise is 2.5 pixels.

Fig. 9d. Sample mean and ± standard deviation of estimated V'_X versus number of frames for experiment 7.

Sample mean and ±σ of V'_Y versus frames

Fig. 9e. Sample mean and ± standard deviation of estimated V'_Y versus number of frames for experiment 7.
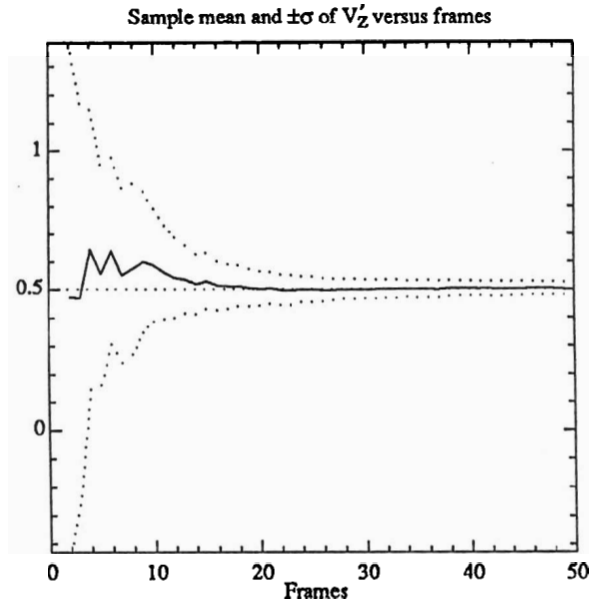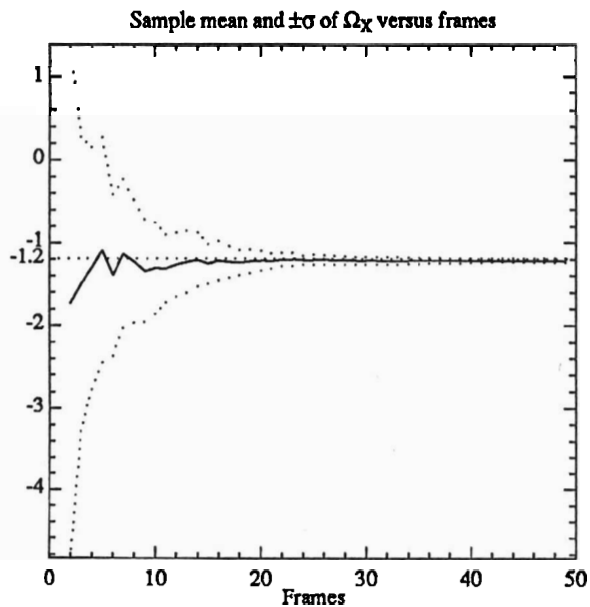
Sample mean and ±σ of V'_Z versus frames

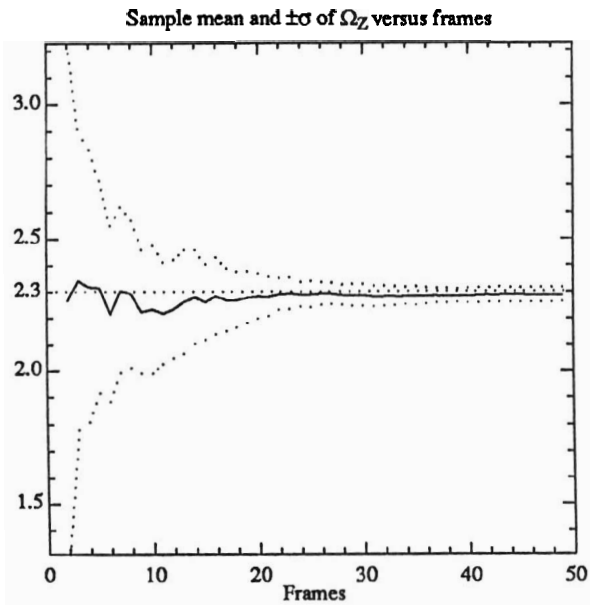Fig. 9f. Sample mean and ± standard deviation of estimated V'_Z versus number of frames for experiment 7.

Sample mean and ±σ of Ω_X versus frames

Fig. 9g. Sample mean and ± standard deviation of estimated Ω_X versus number of frames for experiment 7.

Sample mean and ±σ of Ω_Y versus frames

Fig. 9h. Sample mean and ± standard deviation of estimated Ω_Y versus number of frames for experiment 7.

Sample mean and ±σ of Ω_Z versus frames

Fig. 9i. Sample mean and ± standard deviation of estimated Ω_Z versus number of frames for experiment 7.

Sample mean and ±σ of Z_{10} versus frames
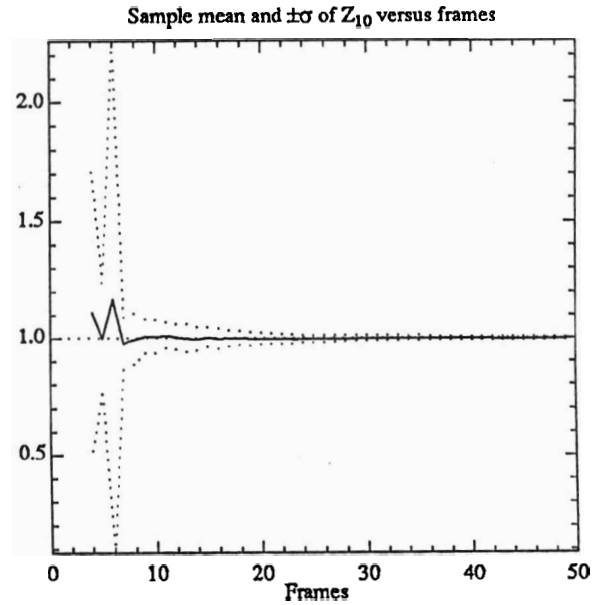

Fig. 9j. Sample mean and ± standard deviation of estimated Z_{10} versus number of frames for experiment 7.
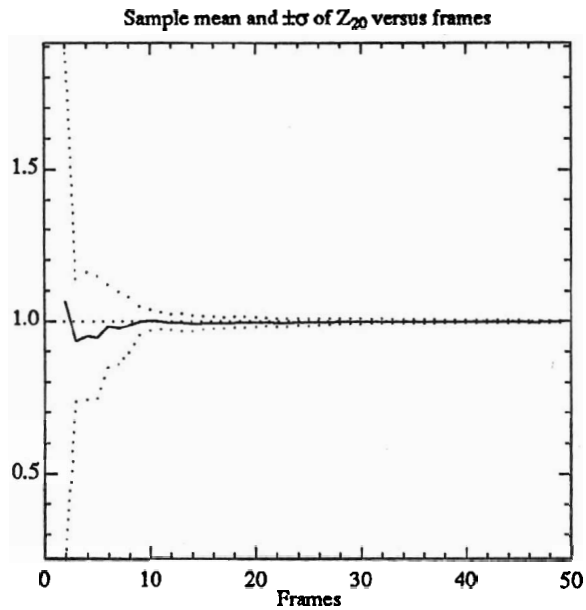
Sample mean and ±σ of Z_{20} versus frames

Fig. 9k. Sample mean and ± standard deviation of estimated Z_{20} versus number of frames for experiment 7.