June 1986 UILU-ENG-86-2218

COORDINATED SCIENCE LABORATORY
College of Engineering

3-D MOTION ESTIMATION, UNDERSTANDING AND PREDICTION FROM NOISY IMAGE SEQUENCES

Juyang Weng
Thomas S. Huang
Narendra Ahuja

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS None
2a. SECURITY CLASSIFICATION AUTHORITY N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) UILU-ENG-86-2218	5. MONITORING ORGANIZATION REPORT NUMBER(S) N/A
6a. NAME OF PERFORMING ORGANIZATION Coordinated Science Laboratory University of Illinois 6b. OFFICE SYMBOL (If applicable) N/A	7a. NAME OF MONITORING ORGANIZATION National Science Foundation
6c. ADDRESS (City, State and ZIP Code) 1101 W. Springfield Ave. Urbana, IL 61801	7b. ADDRESS (City, State and ZIP Code) 1800 G. Street Washington, D.C. 20550
8a. NAME OF FUNDING/SPONSORING ORGANIZATION National Science Foundation 8b. OFFICE SYMBOL (If applicable) N/A	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER ECS-83-52408 ECS-83-19509
8c. ADDRESS (City, State and ZIP Code) 1800 G. Street Washington, D.C. 20550	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO.: N/A; PROJECT NO.: N/A; TASK NO.: N/A; WORK UNIT NO.: N/A
11. TITLE (Include Security Classification) 3-D Motion Estimation, Understanding and Prediction from Noisy Image Sequences	
12. PERSONAL AUTHOR(S) Juyang Weng, Thomas S. Huang, Narendra Ahuja	
13a. TYPE OF REPORT Technical 13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) June 1986 15. PAGE COUNT 58
16. SUPPLEMENTARY NOTATION N/A	
17. COSATI CODES FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) motion, image sequence analysis, dynamic model, motion estimation, motion understanding, motion prediction, computer vision

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This paper presents an approach to understanding general 3-D motion of a rigid body from image sequences. Based on dynamics, a locally constant angular momentum (LCAM) model is introduced. The model is local in the sense that it is applied to a limited number of image frames at a time. Specifically, the model constrains the motion, over a local frame sebsequence, to be a super-position of precession and translation. Thus, the instantaneous rotation axis of the object is allowed to change through the subsequence. The trajectory of the rotation center is approximated by a vector polynomial. The parameters of the model evolve in time so that they can adapt to long term changes in motion characteristics.

The nature and parameters of short term motion can be estimated continuously with the goal of understanding motion through the image sequence. The estimation algorithm presented in this paper is linear. Based on the assumption that the motion is smooth, object positions and motion in the near future can be predicted, and short missing subsequences can be recovered

contd. on back...

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code) 22c. OFFICE SYMBOL None

Noise smoothing is achieved by overdetermination and a least squares criterion. The framework is flexible in the sense that it allows both overdetermination in number of feature points and the number of image frames. The number of frames from which the model is derived can be varied according to the complexity of motion and the noise level so as to obtain stable and good estimates of parameters over the entire image sequence.

Simulation results are given for noisy synthetic data, and images taken of a model airplane.

3-D MOTION ESTIMATION, UNDERSTANDING AND PREDICTION FROM NOISY IMAGE SEQUENCES

JUYANG WENG
THOMAS S. HUANG
NARENDRA AHUJA

Coordinated Science Laboratory
University of Illinois
1101 West Springfield Avenue
Urbana, Illinois 61801
U. S. A.

Index terms :

motion, image sequence analysis, dynamic model, motion estimation, motion understanding, motion prediction, computer vision

3-D MOTION ESTIMATION, UNDERSTANDING AND PREDICTION FROM NOISY IMAGE SEQUENCES

Abstract

This paper presents an approach to understanding general 3-D motion of a rigid body from image sequences. Based on dynamics, a locally constant angular momentum (LCAM) model is introduced. The model is local in the sense that it is applied to a limited number of image frames at a time. Specifically, the model constrains the motion, over a local frame subsequence, to be a superposition of precession and translation. Thus, the instantaneous rotation axis of the object is allowed to change through the subsequence. The trajectory of the rotation center is approximated by a vector polynomial. The parameters of the model evolve in time so that they can adapt to long term changes in motion characteristics.

The nature and parameters of short term motion can be estimated continuously with the goal of understanding motion through the image sequence. The estimation algorithm presented in this paper is linear. Based on the assumption that the motion is smooth, object positions and motion in the near future can be predicted, and short missing subsequences can be recovered.

Noise smoothing is achieved by overdetermination and a least squares criterion. The framework is flexible in the sense that it allows both overdetermination in number of feature points and the number of image frames. The number of frames from which the model is derived can be varied according to the complexity of motion and the noise level so as to obtain stable and good estimates of parameters over the entire image sequence.

Simulation results are given for noisy synthetic data, and images taken of a model airplane.

1. INTRODUCTION

Perception of three-dimensional motion from images is an integral part of vision. It involves estimation of the nature and parameters of 3-D motion, and as a result, prediction of future positions of moving objects. Human vision is adept at using image sequences to understand and predict motion [18]. For example, after a football is kicked off, people can judge whether the football will pass through uprights long before it actually reaches there. In computer vision, cameras must be continuously reoriented to track a moving object for autonomous image acquisition. The motion of a robot arm or a vehicle may have to be estimated and predicted to plan safe motion trajectories. Retrieval and repair of satellites in space requires that the spacecraft rendezvous with the target, which in turn, requires that the spin and the tumbling motion of the target be detected and estimated first. An understanding of the 3-D motion makes it possible to make predictions about future locations and configurations of the moving objects. Such prediction capability allows planning of manipulatory actions on moving objects, e.g., capturing a spacecraft.

We try to characterize quantitatively general 3-D motion from image sequences. The generality of the problem refers to the lack of knowledge about the structure of the objects undergoing motion as well as the type of motion they are undergoing. For example, it may not even be known if the objects are translating, rotating or precessing, much less the motion parameters. Under special restrictions, the problem may be easier to solve although the solution may be of restricted use. Restrictions on both allowed motion as well as object structure have been used to simplify the problem, often making the solution inapplicable to real images. Broida and Chellappa [4] discuss the inference of 2-D motion from 1-D image sequences under the assumption that the object undergoes constant translational and rotational 2-D motion and the structure of the object is known. Yasumoto and Medioni [21] also assume the motion to be constant through the sequence and estimate, through a search in the solution space, the parameters of assumed constant motion from image sequence. In the field of astrodynamics, the dynamic information about he object is required

to be known. For instance, the principal moments of inertia and structures of objects are required [5, 11]. Because the Lagrange equations of rigid body motion are nonlinear [8,14], numerical methods are necessary to solve the dynamics problem [5, 11].

Our goal is to understand the motion with as little *a priori* knowledge as possible. The motion of an object is determined by underlying dynamics. By the analysis of the image sequence under a general dynamic model, the understanding and description of the motion can be derived. Furthermore, based on the motion parameters derived, we can make extrapolations and interpolations through image sequences to predict and recover part of the motion. Clearly, we do not in general know the forces acting on the object and the object structural response to the forces which would otherwise enable us to derive object's 3-D motion from the principles of dynamics. However, it is essential to impose a constraint on the object motion to make the inverse problem of 3-D inference solvable.

In general, the moving objects exhibit a smooth motion, *i.e.*, the motion parameters between consecutive image pairs are correlated. From this assumption and given a sequence of images of a moving rigid object, we determine what kind of local motion the object is undergoing. A *l*ocally *c*onstant *a*ngular *m*omentum model, or LCAM model for short, is introduced. The model assumes short term conservation of angular momentum and a polynomial curve as the trajectory of rotation center. This constraint is the precise statement of what we mean by smoothness of motion. However, we allow the angular momentum, and hence, the motion characteristics of the object to change or evolve over long term. Thus, we do not constrain the object motion by some global model of allowed dynamics.

As a result of the analysis presented in this paper, some of the questions that we can answer are: whether there is precession or tumbling; what the precession is if it exists; how the rotation center of the object (which may be an invisible point !) moves in space; what the future motion would probably be; where a particular object point would be located in image frames or in 3-D at the next several time instants; where the the object would be if it is missing from a image subse-

quence, and what the motion before the given sequence could be.

As a consequence of being able to predict future locations of feature points, only a neighborhood of the predicted position may need to be searched to obtain matching points in successive images.

The imposition of local smoothness of motion constraint helps combat the errors due to noise. One way to combat the effect of such noise would be to use a large number of feature points in the images. However, a large number of feature points is not desirable, especially in the case where very few feature points can be extracted from the objects. The use of image sequences containing a large number of frames is a better way to combat the effect of noise.

Our approach is based on the two-view motion analysis of image sequences consisting of either monocular images, or binocular image pairs. The two-view motion estimation problem is as follows. Given images of a moving object taken at two different time instants, the problem is to estimate the 3-D position transformation of the object between the two time instants. The rotation and translation components of such transformation are referred to as two-view rotation and two-view translation. Generally, they do not represent actual continuous motion undergone by the object between the two time instants. The physical location of the rotation axis is not determined by such two-view position transformation. Two-view motion estimation has been discussed extensively in the literature. Many researchers [6,13,15,16,17,24] have used point correspondences between two image frames to solve this problem. Linear algorithms for two-view motion analysis from point correspondences have been developed by Longuet-Higgins [13], and Tsai and Huang [17]. Line correspondences can also be used to solve the problem (Yen and Huang [22]). An alternative approach is to compute the optical flow field and then estimate motion parameters from optical flow [1,19,20,23]. Zhuang and Haralick [23] give a linear algorithm for such estimation using optical flow. All these motion estimation techniques use monocular images, taken by a monocular sensor such as a single video camera. With such an arrangement the 3-D translation and the range of the object can be determined up to a scale factor. If binocular images are used, we can determine the

absolute translation velocities and ranges of object points. An algorithm to find the rotation and translation given 3-D correspondences has been discussed in Huang and Blostein [10]. An algorithm of Faugeras and Hebert [7] can be used to find the least squares solution of the motion parameters in the presence of noise. A review of recent results in motion analysis is given in [9].

The approach presented in this paper can use either feature points or optical flow to solve two-view motion parameters. We use feature points in the discussion here. We assume that there is a single rigid object in motion, the correspondences of points between images are given, and the motion does not exhibit any discontinuities such as those caused by collisions.

In Section 2 we first present the LCAM model based on dynamics. Then the solutions of the model parameters are discussed and the relationship between continuous motion and discrete two-view motion is described. The approach to estimating these parameters in the presence of noise is discussed in Section 3. Section 3 also deals with the local understanding, prediction and recovering of the motion. Some particular properties of monocular vision are discussed in Section 4. Section 5 gives the results of simulations. Section 6 presents a summary.

2. THE LCAM MODEL

This section consists of four subsections. Subsection 2.1 deals with the general motion of a rigid body in 3-D. Subsection 2.2 is devoted to the motion of the rotation center. The trajectory of the rotation center is approximated by a vector polynomial as a function of time. Subsection 2.3 discusses the solution of these coefficient equations. The relationship between the continuous precession of the LCAM model and the discrete two-view motion is investigated in Subsection 2.4.

2.1 Motion of a Rigid Body in 3-D

All external forces acting on a body can be reduced to a total force **F** acting on a suitable point Q, and a total applied torquer **N** about Q. For a body moving freely in space, the center of mass is to be taken as the point Q. If the body is constrained to rotate about a fixed point, then that point is to be taken as the point Q. This point may move with the supports. Let m be the mass of the body. the motion of the center of mass is given by

$$\mathbf{F} = \frac{d}{dt}(m\,\mathbf{V}) \tag{1}$$

Let **L** be the angular momentum of the body. The torque **N** and the angular momentum **L** satisfy [8,14] :

$$\mathbf{N} = \frac{d\,\mathbf{L}}{dt} \tag{2}$$

The rotation is about the point Q, which will be referred to as the rotation center. In the remainder of this subsection, we concentrate on the rotation part of the motion. The motion of the center of mass will be discussed in the next subsection.

In matrix notation, the angular momentum **L** can be represented by

$$L = I\,\omega$$

or writing in components:

$$\begin{bmatrix} L_x \\ L_y \\ L_z \end{bmatrix} = \begin{bmatrix} I_{xx} & I_{yx} & I_{zx} \\ I_{xy} & I_{yy} & I_{zy} \\ I_{xz} & I_{yz} & I_{zz} \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}$$

where

$$I_{xx} = \int (y^2+z^2)dm \quad I_{yy} = \int (z^2+x^2)dm \quad I_{zz} = \int (x^2+y^2)dm$$

$$I_{zx} = I_{xz} = -\int xz \ dm \quad I_{yx} = I_{xy} = -\int xy \ dm \quad I_{zy} = I_{yz} = -\int zy \ dm$$

and where ω is angular velocity. The above integrals are over the mass of the body.

If the coordinate axes are the principal axes of the body [8.14], the inertia tensor I takes the diagonal form:

$$I = \begin{bmatrix} I_{xx} & 0 & 0 \\ 0 & I_{yy} & 0 \\ 0 & 0 & I_{zz} \end{bmatrix}$$

Referred to a coordinate system fixed on such a rotating body, (2) becomes

$$\begin{aligned} N_x &= I_{xx}\dot{\omega}_x + \omega_y \omega_z (I_{zz} - I_{yy}) \\ N_y &= I_{yy}\dot{\omega}_y + \omega_z \omega_x (I_{xx} - I_{zz}) \\ N_z &= I_{zz}\dot{\omega}_z + \omega_x \omega_y (I_{yy} - I_{xx}) \end{aligned} \tag{3}$$

where $(N_x, N_y, N_z) = N$. These are known as Euler's equations for the motion of a rigid body. These equations are nonlinear and have generally no closed-form solutions. Numerical methods are needed to solve them.

Clearly the motion of a rigid body under external forces is complicated. In fact even under no external forces, the motion remains to be complex. Perspective projection adds further complexity to the motion as observed in the image. However, in a short time interval, realistic simplifications can be introduced. One simplification occurs if we ignore the impact of the external torque over short time intervals. If there is no external torque over a short time, there is no change

in the angular momentum of the object. Thus, if we have a dense temporal sequence of images, we can perform motion analysis over a small number of successive frames under the assumption of locally constant angular momentum. Another simplification occurs if the body possesses an axis of symmetry. The symmetry here means that at least two of I_{xx}, I_{yy}, I_{zz} are equal. Cylinders and disks are such examples. Most satellites are also symmetrical or almost symmetrical in this sense.

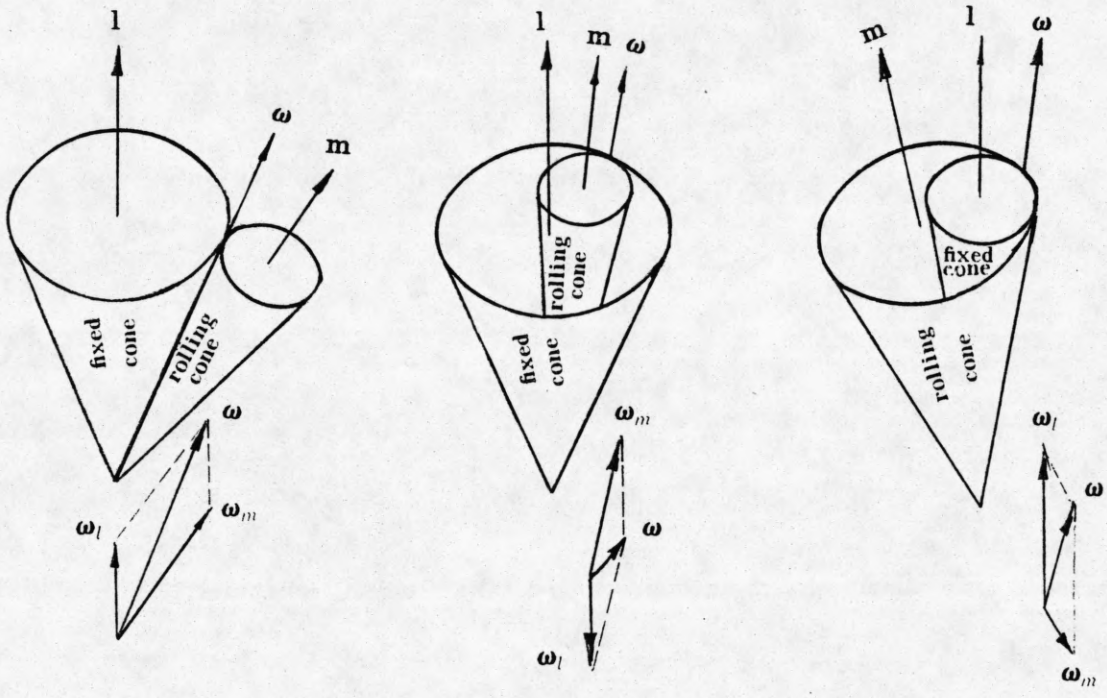

Fig.1 The precessional motion of a torque-free symmetrical rigid body

Under the above two simplifications, Euler's equations are integrable [8,14]. The motion is such that the body rotates about its axis of symmetry **m**, at the same time it rotates about a spatially fixed axis **l**. The motion can be represented by a rotating cone that rolls along the surface of a fixed cone without slipping as shown in Fig.1, where the body is fixed on the rolling cone, the axis of symmetry coincides with that of rolling cone, and the center of mass or the fixed point Q of the body coincides with the apices of the cones. Then, the motion of the rolling cone is the same as the motion of the body. Fig.1 gives three possible configurations of the rolling cone and the fixed cone.

Let $\boldsymbol{\omega}_l$ be the angular velocity at which the rolling cone rotates about $\mathbf{l}$. Let $\boldsymbol{\omega}_m$ be the angular velocity at which the rolling cone rotates about its own axis of symmetry $\mathbf{m}$. Then the instantaneous angular velocity $\boldsymbol{\omega}$ is the vector sum of $\boldsymbol{\omega}_l$ and $\boldsymbol{\omega}_m$ as shown in Fig.1. The magnitudes of $\boldsymbol{\omega}_m$ and $\boldsymbol{\omega}_l$ are constant. So the magnitude of the instantaneous angular velocity is also constant. This kind of motion about a point is called precession in the following sections and it represents the restriction imposed by our model on the allowed object rotation.

A special case occurs when $\mathbf{m}$ is parallel to $\mathbf{l}$. Then $\boldsymbol{\omega}$ is also parallel to $\mathbf{l}$. So the instantaneous rotation axis does not change its orientation in motion. This type of motion is called motion without precession.

2.2 Motion of Rotation Center

In the rest of this paper column vectors and matrices will be used very often and will be generally denoted by italics. Column vectors will also denote column matrices. So dot and cross operations as well as matrix multiplications will be applied to column vectors. F_i will denote an image frame taken at time t_i, $i = 0, 1, \cdots$.

The location of rotation center is changed with time. Assume the trajectory of the rotation center is smooth, or specifically, it can be expanded into a Taylor series:

$$Q(t) = Q(t_0) + \frac{1}{1!}Q'(0)(t - t_0) + \frac{1}{2!}Q''(0)(t - t_0)^2 + \cdots \tag{4}$$

If the time intervals between image frames are short, we can estimate the trajectory by the first k terms. We get a polynomial of time t. The coefficients of the polynomial are 3-vectors. Letting $\frac{1}{(i-1)!}Q^{(i-1)}(0) = b_i$, $i = 1,2,3,\cdots$, we have

$$Q_i = b_1 + b_2(t_i - t_0) + b_3(t_i - t_0)^2 + \cdots\cdots + b_k(t_i - t_0)^{k-1} \tag{5}$$

For simplicity, we assume the time intervals between image frames are constant c, *i.e.*, $t_i = ci + t_0$. From (5) we get

$$Q_i = b_1 + cb_2 i + c^2 b_3 i^2 + \cdots + c^{k-1} b_k i^{k-1} \tag{6}$$

Letting $a_j = c^{j-1} b_j$, $j = 1,2,\cdots$, we get

$$Q_i = a_1 + a_2 i + a_3 i^2 + \cdots + a_k i^{k-1} \tag{7}$$

Equation (7) is the model for the motion of rotation center. The basic assumption we made is that the trajectory can be approximated by a polynomial. If the motion is smooth and the time interval covered by the model is relatively short, Equation (7) is a good approximation of the trajectory.

Together with the precession model presented in the previous subsection, we have the complete LCAM model. Therefore, the model is characterized by locally constant angular momentum,

i.e., the angular momentum of the moving object can be treated as constant over short time intervals. Though we derive this model from the assumption of constant angular momentum and object symmetry, the condition leading to such motion is not unique. In other words, the motion model we derived applies to any moving objects whose rotation can be locally modeled by such motion: the rotation about a fixed-on-body axis that rotates about a spatially fixed axis, and whose translation can be locally modeled by a vector polynomial.

Our goal here is to understand 3-D motion of an object over an extended time period using the two-view motion analysis of images taken at consecutive time instants. Thus we would first estimate the two-view motion parameters of the moving object.

The image sequence can be either monocular or binocular. In the binocular case, at each time instant we take a pair of images using two cameras in certain configuration. From one such image pair, we can find the 3-D coordinates of a point assuming its locations in the two images are known. The 3-D coordinates of an object point at two time instants define a point correspondence. At least three point correspondences are needed to uniquely determine the relative displacement of a rigid body between these two time instants. The displacement can be represented by a rotation about an axis located at the origin of a world coordinate system, and a translation [2]. We will call this displacement as two-view motion. In the monocular case, only one camera is used. At each time instant, one image is taken which is a perspective projection of the object at that time instant. The image coordinates of an object point at two time instants define a point correspondence in the monocular case. At least eight point correspondences are needed to uniquely determine rotation and translation direction of two-view motion using linear algorithms [17,24]. The magnitude of translation vector can not be determined generally from monocular images. More point correspondences are needed to improve accuracy in the presence of noise.

Let the column vector P_0 be the 3-D coordinates of an object point at time t_0. P_1 be that of the same point at time t_1, R_1 be the rotation matrix from time t_0 to t_1, and T_1 be the corresponding translation vector. Then, P_0 and P_1 are related by

$$P_1=R_1P_0+T_1 \tag{8}$$

where R_1 represents a rotation about an axis through the origin.

Given a set of point correspondences, R_1 and T_1 can be determined by two-view motion analysis. In the case of monocular vision, the translation vector can only be determined up to a positive scale factor. *i.e.* only the direction of T, $\hat{T}=T/\|T\|$, can be determined from the perspective projection.

In equation (8) letting P_0 be at the origin, it is clear that T_1 is just the translation of the point at origin. For any point Q_0, we can translate the rotation axis so that it goes through Q_0 and rotate P_0 about the axis at the new location. Mathematically, from (8) we have

$$P_1=R_1(P_0-Q_0)+(R_1Q_0+T_1) \tag{9}$$

Compared with (8), (9) tells us that the same motion can be represented by rotating P_0 about Q_0 by R_1, and then translating by $R_1Q_0+T_1$. Because Q_0 is arbitrarily chosen, there are infinitely many ways to select the location of the rotation axis. This is an ambiguity problem in motion understanding from image sequences. If we let the rotation axis always be located at origin, the trajectory described by R_i and T_i $i=1,2,3\cdots$ would be like what is showed in Fig.2, which is very unnatural.

In Fig.2 the real trajectory of the center of the body is the dashed line. However, neither the rotation nor the translation components show this trajectory. As we discussed in Subsection 2.1, the center of mass of a body in free motion satisfies Newton's equation of motion of a particle (1). Rotation is about the center of mass (or fixed point if it exists). So motion should be expressed in two parts, the motion of the rotation center (the center of mass or the fixed point), and the rotation about the rotation center.

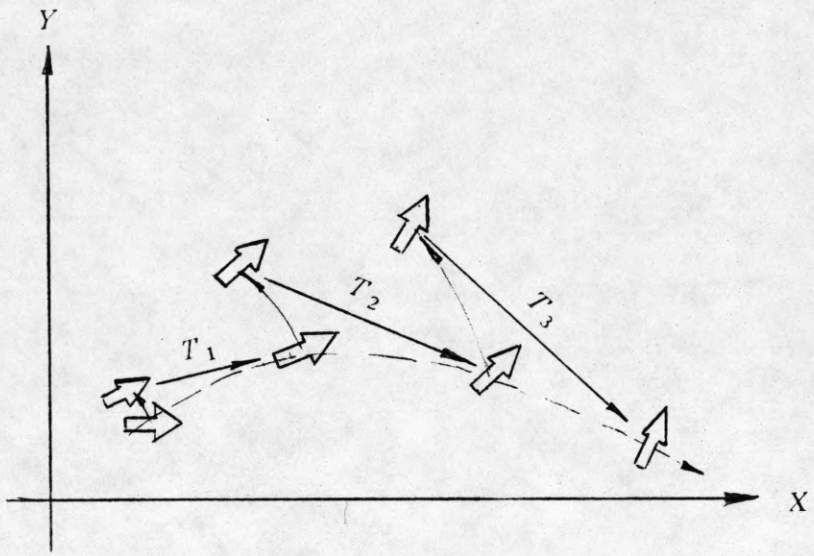

Fig.2 Trajectory described by R_i and T_i
if rotation axis is always located at origin

Let Q_i be the position vector of the rotation center at time t_i, P_i be another point at time t_i, R_i be the rotation matrix from t_{i-1} to t_i, T_i be the translation vector from t_{i-1} to t_i. From (8) we have

$$Q_1=R_1Q_0+T_1 \tag{10}$$

Substituting (10) into (9) we get

$$P_1=R_1(P_0-Q_0)+Q_1 \tag{11}$$

From (10) we have

$$-R_1Q_0+Q_1=T_1$$

Similarly we get equations for motion from t_{i-1} to t_i, $i=1,2,\cdots,f$:

$$-R_1Q_0+Q_1=T_1$$

$$-R_2Q_1+Q_2=T_2 \tag{12}$$

.....

$$-R_f\,Q_{f-1}+Q_f=T_f$$

Equations (12) give the relationship among locations of rotation center, the two-view rotation matrices and two-view translation vectors. The equations (12) will be referred to as center motion equations.

Substituting (7) into (12), we get

$$\begin{aligned}
&(I-R_1)a_1+a_2+a_3+\cdots+a_k=T_1\\
&(I-R_2)a_1+(2I-R_2)a_2+(4I-R_2)a_3+\cdots+(2^{k-1}I-R_2)a_k=T_2\\
&\cdots\cdots\\
&(I-R_f)a_1+(fI-(f-1)R_f)a_2+(f^2I-(f-1)^2R_f)a_3+\cdots+(f^{k-1}I-(f-1)^{k-1}R_f)a_k=T_f
\end{aligned} \tag{13}$$

Vector equations (13) are referred to as the coefficient equations. Both sides of the equations are 3-vectors. There are f equations in k unknown 3-vectors. Let $A=(a_1, a_2, \cdots, a_k)^t$, $T=(T_1, T_2, \cdots, T_f)^t$, D be the coefficient matrix of the unknowns in (13). The element of D at i-th row and j-th column is d_{ij}, *i.e.*, $D=[d_{ij}]_{f\times k}$. We have

$$d_{ij}=j^{j-1}I-(i-1)^{j-1}R_i$$

We can rewrite the coefficient equations (13) as

$$DA=T \tag{14}$$

D and T are determined by two-view motion analysis. The problem here is to find A, the coefficients of the polynomial in (7).

2.3 Solutions of Coefficient Equation

Let $f = k$, then the matrix D is a square matrix. We wish to know whether the linear equations (14) have a solution. If a solution exists, is it unique ? If it is not unique, what is the general solution ?

The solution of the coefficient equations depends on the types of motion, or the rotation matrices R_i and the translation vectors T_i. Let us first consider a simpler case, where $k = 2$. This means that the trajectory of the rotation center is locally approximated by a motion of constant velocity. Three frames are used in this case. The coefficient equations become

$$(I - R_1)a_1 + a_2 = T_1 \tag{15.1}$$

$$(I - R_2)a_1 + (2I - R_2)a_2 = T_2 \tag{15.2}$$

Solving for a_2 in (15.1) and substituting it into (15.2), we get

$$(I - 2R_1 + R_2R_1)a_1 = (2I - R_2)T_1 - T_2 \tag{16}$$

If $I - 2R_1 + R_2R_1$ is nonsingular, a_1 can be uniquely determined from (16) :

$$a_1 = (I - 2R_1 + R_2R_1)^{-1}((2I - R_2)T_1 - T_2) \tag{17}$$

Then a_2 is determined from (15.1) :

$$a_2 = T_1 - (I - R_1)a_1$$

Appendix 3 shows that $(I - 2R_1 + R_2R_1)$ is nonsingular if and only if the following two conditions are both satisfied: 1) the axes of rotations, represented by R_1 and R_2, respectively, are not parallel. 2) Neither rotation angle is zero. Condition 2) is usually satisfied if the motion is not pure translation. If condition 1) is not satisfied, the solution of equations (15) is not unique and have some structure. To show this, assume the rotation axes of R_1 and R_2 are parallel. Let w be any vector parallel to these axes. Because any point on the rotation axis remains unchanged after rotation, we have $R_1w = w$, $R_2w = w$. For any solution a_1 and a_2, $a_1 + cw$ and a_2 is another solution,

where

c is an arbitrary real constant. So there exist infinitely many solutions.

The following theorem presents the results for the general case.

THEOREM 1. In coefficient equations, let f=k. Define S_k^j to be a 3 by 3 matrix :

$$S_k^j=\sum_{l=0}^{k-j}(-1)^l\binom{k}{l}R_{k-l}R_{k-l-1}\cdots R_{j+1}I \qquad j=0, 1, 2, \cdots, k$$

Define number u_{ij} :

$$u_{ij}=\sum_{m=1}^{i}(-1)^{i-m}m^{j-1}\binom{i}{m} \qquad j=i+1, i+2, \cdots, k$$

Then

$$S_k^0 a_1=-\sum_{l=1}^{k}S_k^l T_l$$

$$a_k=\frac{-1}{(k-1)!}\left(\sum_{m=1}^{k-1}S_{k-1}^m T_m+S_{k-1}^0 a_1\right)$$

$$a_{k-1}=\frac{-1}{(k-2)!}\left(\sum_{m=1}^{k-2}S_{k-2}^m T_m+S_{k-2}^0 a_1+u_{k-2\,k}a_k\right)$$

$$a_{k-2}=\frac{-1}{(k-3)!}\left(\sum_{m=1}^{k-3}S_{k-3}^m T_m+S_{k-3}^0 a_1+u_{k-3\,k}a_k+u_{k-3\,k-1}a_{k-1}\right)$$

.....

$$a_3=\frac{-1}{2!}\left(\sum_{m=1}^{2}S_2^m T_m+S_2^0 a_1+u_{2\,k}a_k+u_{2\,k-1}a_{k-1}+\cdots+u_{2\,4}a_4\right)$$

$$a_2=\frac{-1}{1!}\left(\sum_{m=1}^{1}S_1^1 T_1+S_1^0 a_1+u_{1\,k}a_k+u_{1\,k-1}a_{k-1}+\cdots+u_{1\,3}a_3\right)$$

Proof. See Appendix 1.

If S_k^0 is not singular, the first equation given by Theorem 1 uniquely determines a_1. Then a_k, a_{k-1}, $\cdots$, a_2 can be determined, sequentially, by the second, third, ..., and last equations in Theorem 1. So if S_k^0 is not singular, solution is unique.

THEOREM 2. In the case of rotation without precession, let w be any column vector parallel to the rotation axes. Then,

$$S_k^0 w = 0 \tag{18}$$

and for any vector a

$$(S_k^0 a) \cdot w = 0 \tag{19}$$

Proof. See Appendix 2.

From Theorem 1, we have

$$S_k^0 a_1 = -\sum_{l=1}^{k} S_k^l T_l \tag{20}$$

In the case of rotation without precession, equation (18) implies S_k^0 is singular. From (19), the left-hand side of (20) is orthogonal to w. However if the real trajectory of the rotation center is not exactly a j-th degree polynomial with $j \leqslant k-1$ in (7), the right-hand side of (20) can be any vector, which may not be orthogonal to w. This means that no solution exists for equation (20). If the real trajectory is a j-th degree polynomial with $j \leqslant k-1$, then equation (20) has a solution by our derivation of (20). Since equation (7) is usually only an approximation of the real trajectory, a least squares solution of (20) can serve our purpose. Let $\hat{a}_1$ be the least squares solution of (20) which is solved by using independent columns of S_k^0. If the rank of S_k^0 is 2, which is generally true for motion without precession, the general solution is then $a_1 = \hat{a}_1 + cw$, where c is any real number. All general solutions $\{\hat{a}_1 + cw\}$ form a straight line in 3-D space. From equation (7), this line gives the location and direction of the two-view rotation axis of the motion between time instants t_0 and t_1. From Theorem 2 we have $S_{k-1}^0 w = 0$, $S_{k-2}^0 w = 0$, $\cdots$, $S_1^0 w = 0$. Then $S_{k-1}^0 a_1 = S_{k-1}^0 \hat{a}_1$, $S_{k-2}^0 a_1 = S_{k-2}^0 \hat{a}_1$, $\cdots$, $S_1^0 a_1 = S_1^0 \hat{a}_1$. By the equations given by Theorem 1, the unknowns a_k, a_{k-1}, $\cdots$, a_2 are determined without knowing the undetermined scale c.

If the motion is pure translation without rotation, all the rotation matrices R_i, $i = 1, 2, \cdots, k$, are unit matrix I. S_k^0 is zero matrix. The first three columns of D are zero. a_1 can not be determined by coefficient equations. From Theorem 1, $a_2, a_3, \cdots, a_k$, can still be determined by coefficient

equations. Because no rotation exists, any point can be considered as a rotation center. Equation (7) can be used to approximate the trajectory of any object points.

Thus the solutions of the coefficient equations can be summarized as follows.

1. In the case of rotation with precession, the solution of the coefficient equations is generally unique. The trajectory of the rotation center is described by (7).

2. In the case of rotation without precession, the general solution of a_1 gives the two-view rotation axis of the first two-view motion. All other coefficients $a_2, a_3, \cdots, a_k$ are generally determined uniquely by Theorem 1. So the two-view rotation axes of all two-view motions are determined by (7). Because no precession exists, any point on the rotation axis can be considered as the rotation center. This is the meaning of the general solution a_1. Once a particular point on the rotation axis is chosen as the rotation center, its trajectory is described by equation (7). Fig.3 shows the possible "parallel" trajectories of the rotation center depending on which point on the axis is chosen as the rotation center.

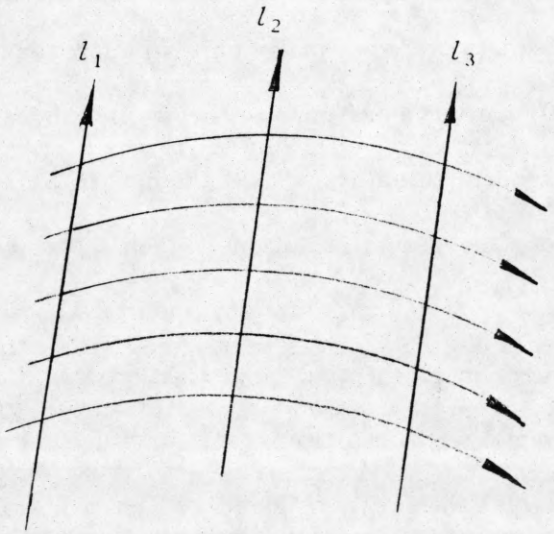

Fig.3 The possible trajectories of rotation centers when rotation axes are parallel.

3. In the case of pure translation without rotation, $a_2, a_3, \cdots, a_k$ can still be determined by coefficient equations. However a_1 can not be determined by coefficient equations. a_1 can be chosen to be the position of any object point at time t_0. Then equation (7) describes the trajectory of this point.

In the presence of noise, both a large number of point correspondences and a large number of image frames provide overdetermination. Faugeras and Hebert [7] present an algorithm which can be used for the least squares estimation of two-view motion parameters. To use overdetermination based on a large number of frames, we let $f > k$ in the coefficient equations (13). In fact, the coefficient matrix S_k^0 is essentially a high order deference. This is shown in Lemma 1 of Appendix 1. S_k^0 tends to be ill-conditioned when k gets large. This means $f > k$ is more important when k is large. If $f > k$, equation (14) can be solved by a least squares method. We find a solution A such that

$$\| DA - T \| = min \tag{21}$$

In the case of motion with precession, all the columns of D are generally independent. The least squares solution is

$$A = (D^t D)^{-1} DT \tag{22}$$

In the case of motion without precession, the column vectors of D are linearly dependent. This can be shown by letting a_1 in equation (13) be a non-zero vector parallel to the two-view rotation axes. Then the first three columns of D linearly combined by a_1 is zero vector. To get the least squares solution of the coefficient equations (13), the largest set of independent columns of D should be found or tolerance-based column pivoting should be made [12]. Theorem 1 solves $a_2, a_3, \cdots, a_k$. This means the last $3k-3$ columns of D are always independent. In the presence of noise the columns of D are very unlikely to be exactly linearly dependent even in the case of motion without precession.

2.4 Continuous Precession and Discrete Two-View Motion

The LCAM model we discussed is based on continuous precessional motion. We must find the relationship between continuous precession and two-view motion, before we can estimate the precession parameters of our model based on discrete two-view motions.

As we discussed in Subsection 2.1, a precession can be considered as a rolling cone which rolls without slipping upon a fixed cone. The angular frequency at which the symmetrical axis of the rolling cone rotates about the fixed cone is constant.

Assuming at time t_1, an edge point A' on the rolling cone touches an edge point A on the fixed cone as shown in Fig 4 . After a certain amount of rolling, the touching points become B' on the rolling cone and B on the fixed cone. Let θ be the central angle of points A' and B', and ϕ be that of A and B. Let r and r' be the radii of circles O and O', respectively. The arc length between A and B is equal to that between A' and B'. So $\phi r = \theta r'$ or $\phi \sin\alpha = \theta \sin\beta$, where α and β are generating angles of the fixed cone and the rolling cone respectively. We get

$$\frac{\theta}{\phi} = \frac{\sin\alpha}{\sin\beta} \tag{23}$$

The precession consists of two rotational components. One is the rotation of the rolling cone about its own symmetrical axis. The other is the rotation of the rolling cone about the fixed cone. From Fig.4 it can be readily seen that the relative position of the rolling cone and the fixed cone is uniquely determined if the touching points of the two cones are determined. Or alternatively, starting from the previous position, the new position of the rolling cone is determined if the two angles ϕ and θ are determined. So no matter how we order these two rotational components, the final positions are identical as long as the angle ϕ and θ are kept unchanged. We can first rotate the rolling cone about its axis m and then rotate the rolling cone about l, or vice versa.

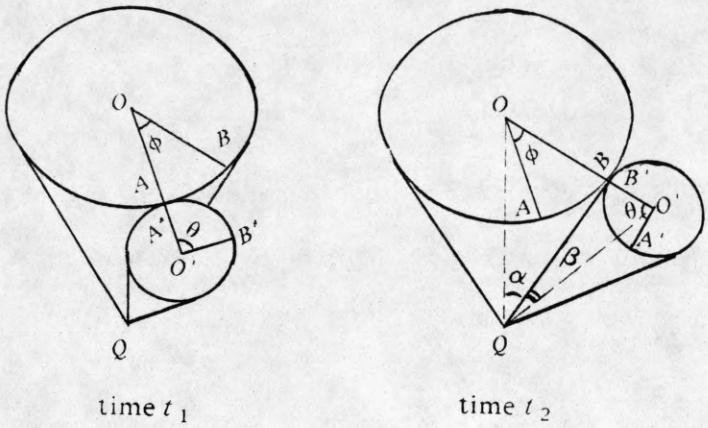

Fig.4 The relation between rotation angles θ and ϕ

We hope to find the equivalent two-view rotation axis of this continuous motion between two frames at time t_1 and time t_2, respectively in Fig.4. If we can find two fixed points which stay in the same positions before and after the motion, then the two-view rotation axis must go through these points. One trivial fixed point is the apex Q of the cones. Another fixed point can be founded as follows: In Fig.5 let the midpoint of arc AB touch the rolling cone (at time $(t_1+t_2)/2$). Extend line OB so that it intersects the plane containing Q, O' and B' at a point P_1. Extend line OA so that it intersects the plane containing Q, O' and A' at a point P_2. Draw a circle centered at O and passing through P_1 and P_2. Then the midpoint P of arc P_1P_2 is a fixed point. This can be seen by noting that the rolling cone can also reach its position at next time instant t_2 in an alternative manner as follows. First, rotate the rolling cone with slipping about l by angle $\phi/2$, thus rotating P to its new position at P_1, and axis m reaches the position shown in Fig.5. Then rotate the rolling cone with slipping about its own axis m by angle θ. Point P now reaches position P_2. Finally, rotate the rolling cone with slipping about l again by angle $\phi/2$, taking the rolling cone to the position at time instant t_2. This takes the point P back to its starting position. So the two-view

rotation axis w found by two-view motion analysis from two image frames, goes through Q and P. Notice that the angular frequency, at which the symmetrical axis of the rolling cone rotates about the fixed cone is constant. From the way of finding P, it is clear that the two-view rotation axis also rotates about l by a constant angle between consecutive frames. So we have the following theorem:

THEOREM 3: For a body undergoing the precessional motion of the LCAM model, the two-view rotation axis between constant time intervals changes by rotating about the precession vector by a constant angle.

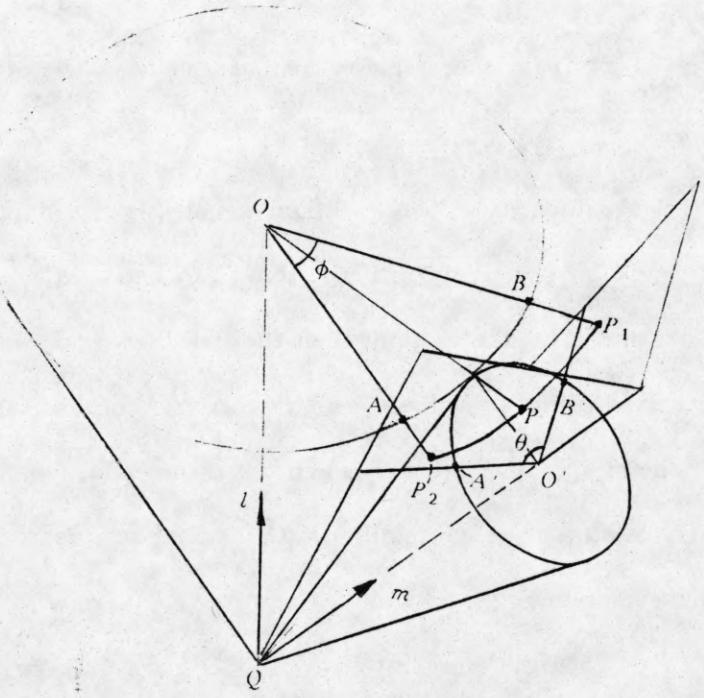

Fig.5 Finding fixed points for two-view rotation

Without loss of generality, we assume the time intervals between consecutive image frames are of unit length. We define the precession vector to be a unit vector l parallel to the symmetrical axis of the fixed cone, define the precession angular frequence ϕ to be the angular frequency at

which the symmetrical axis of rolling cone rotates about precession axis, define the i-th body vector m_i to be a unit vector parallel to the symmetrical axis of the rolling cone at time t_i, and define the body rotation angular frequency θ to be the angular frequency at which the rolling cone rotates about its symmetrical axis.

From image sequences we find estimates of two-view motion parameters. They are the i-th two-view rotation axis vector, n_i, a unit vector parallel to the two-view rotation axis between time instants t_{i-1} and t_i; the corresponding i-th rotation angle ψ_i and i-th translation vector T_i. Fig.6 shows the parameters of continuous motion and discrete two-view motion.

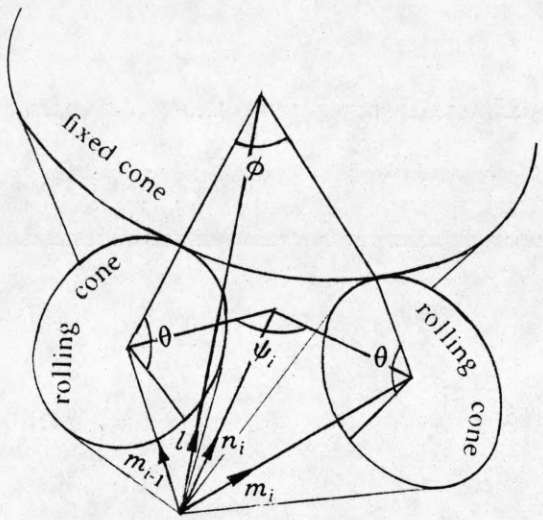

Fig.6 Parameters of continuous precession and discrete two-view motion

Let $R(n,\theta)=[r_{ij}]$ denote the rotation matrix representing a rotation with axis unit vector $n=(n_x, n_y, n_z)$ and rotation angle θ. *i.e.* [2]:

$$R(n,\theta)=\begin{bmatrix} (n_x^2-1)(1-\cos\theta)+1 & n_x n_y(1-\cos\theta)-n_z\sin\theta & n_x n_z(1-\cos\theta)+n_y\sin\theta \\ n_y n_x(1-\cos\theta)+n_z\sin\theta & (n_y^2-1)(1-\cos\theta)+1 & n_y n_z(1-\cos\theta)-n_x\sin\theta \\ n_z n_x(1-\cos\theta)-n_y\sin\theta & n_z n_y(1-\cos\theta)+n_x\sin\theta & (n_z^2-1)(1-\cos\theta)+1 \end{bmatrix} \quad (24)$$

THEOREM 4. The continuous precession parameters and discrete two-view motion parameters are

related by

$$R(l,\phi)R(m_{i-1},\theta)=R(n_i,\psi_i) \tag{25}$$

$$R(m_i,\theta)R(l,\phi)=R(n_i,\psi_i) \tag{26}$$

Proof. From time t_{i-1}, to time t_i, the body moves from its previous position to a new position. From Fig. 6 the new position of the rolling cone (or the body) can be reached in the following way: First, the rolling cone rotates about its body vector m_{i-1} by angle θ. Then, the rolling cone rotates about the precession vector l by angle ϕ. The two-view motion combines these two motions into one, which is the rotation about the two-view rotation axis vector n_i by angle ψ_i. We get equation (25). Similarly if we change the order of these two rotational components we get equation (26).

From Theorem 3, the two-view rotation axis rotates about the precession vector. So the precession vector l is perpendicular to n_i-n_{i-1} and $n_{i-1}-n_{i-2}$. The sign of l is arbitrary. So l can be determined by

$$l=\frac{(n_i-n_{i-1})\times(n_{i-1}-n_{i-2})}{\|(n_i-n_{i-1})\times(n_{i-1}-n_{i-2})\|} \tag{27}$$

The precession angular frequency ϕ can be determined by finding the angle by which n_{i-1} rotates about l to reach n_i. The projections of n_{i-1} and n_i onto a plane that is perpendicular to l are $n_{i-1}-(n_{i-1}\cdot l)l$ and $n_i-(n_i\cdot l)l$, respectively. The angle between these two projections gives the absolute value of ϕ:

$$|\phi|=\cos^{-1}\frac{(n_{i-1}-(n_{i-1}\cdot l)l)\cdot(n_i-(n_i\cdot l)l)}{\|n_{i-1}-(n_{i-1}\cdot l)l\|\,\|n_i-(n_i\cdot l)l\|} \tag{28}$$

The sign of ϕ is the same as the sign

$$(n_{i-1}\times n_i)\cdot l \tag{29}$$

After l and ϕ are found by equations (27), (28) and (29), $R(l,\phi)$ can be calculated by (24). $R(m_{i-1},\theta)$ and $R(m_i,\theta)$ can be determined by (25) and (26):

$$R(m_{i-1}, \theta)=R^{-1}(l, \phi)R(n_i, \psi_i) \tag{30}$$

$$R(m_i, \theta)=R(n_i, \psi_i)R^{-1}(l, \phi) \tag{31}$$

We can determine m_{i-1}, m_i and θ by (30) and (31), because n and θ can be determined from $R(n, \theta)$ [2]:

$$\theta=\pm\cos^{-1}\left(\frac{r_{11}+r_{22}+r_{33}-1}{2}\right) \tag{32}$$

$$l=\frac{(r_{32}-r_{23}, r_{13}-r_{31}, r_{21}-r_{12})^t}{\|(r_{32}-r_{23}, r_{13}-r_{31}, r_{21}-r_{12})\|} \tag{33}$$

So we get the following

THEOREM 5. The precession vector, precession angular frequency, body axes and body rotation angular frequency which define the precession part of the LCAM model can all be determined from three consecutive two-view motions, or four consecutive image frames.

In addition to these basic parameters which uniquely determine the motion of the model, some other parameters can also be determined from these basic parameters. For example, the generating angles α and β of the fixed cone and the rolling cone, respectively, in Fig.4 can also be determined from l, ϕ, m_i, θ and equation (23).

3. ESTIMATION, UNDERSTANDING AND PREDICTION

The LCAM model is applied to subsequences recursively. The parameters of the model describe the current local motion. The following questions can be answered. Is there precession? If so, what are the precession parameters? What are the current or previous body vectors? What is the body rotation angular frequency? What is the probable motion for the next several time intervals? What are the probable locations of the feature points at the next several time instants? If the moving object were occluded in some of the previous image frames, what are the motion and the locations of these feature points during that time period?

The number of frames covered by a LCAM model can be made adaptive to the current motion. The number can be changed continuously to cover as many frames as possible so long as the constant angular momentum assumption is approximately true during the time period to be covered. The value of the number of frames chosen can be based on the accuracy with which the model describes the current set of consecutive frames. The residuals of least squares solutions and the variances of the model parameter samples indicate the accuracy. The noise level also affects the residuals and the variances of parameter samples. However, the noise level is relatively constant or can be measured. The resolution of cameras and the viewing angle covering the object generally determine the noise level. The noise can be smoothed by determining the best time intervals and the number of frames covered by the model, according to the current motion. Because the LCAM model is relatively general, the time interval a LCAM model can cover is expected to be relatively long in most cases.

The following part deals with the estimation of model parameters using overdetermination.

After finding two-view rotation axis vectors $n_1, n_2, \cdots, n_f$, precession vector l should be orthogonal to $n_2-n_1, n_3-n_2, \cdots, n_f-n_{f-1}$. However, because of noise, this may not be true. So we find l such that the sum of squares of the projections of l onto $n_2-n_1, n_3-n_2, \cdots, n_f-n_{f-1}$ is the smallest. Let

$$A = \begin{bmatrix} (n_2 - n_1)^t \\ (n_3 - n_2)^t \\ \cdots\cdots \\ (n_f - n_{f-1})^t \end{bmatrix} \tag{34}$$

We are to find unit vector l, such that

$$\| Al \| = \min \tag{35}$$

Or equivalently,

$$l^t A^t A l = \min \tag{36}$$

The solution l of (36) is the unit eigenvector corresponding to the smallest eigenvalue of $A^t A$ [3].

Let the precession angular frequency determined from (28) and (29) be ϕ_i. The precession angular frequency of the model ϕ can be estimated by the mean

$$\phi = \frac{1}{f-1} \sum_{i=2}^{f} \phi_i \tag{37}$$

Let the body rotation angular frequency determined by (31) be θ_i. Body rotation angular frequency of the model can be estimated by mean

$$\theta = \frac{1}{f} \sum_{i=1}^{f} \theta_i \tag{38}$$

Body vectors are estimated by averaging two consecutive two-view motion using (30) and (31), respectively.

Two-view angular frequency can also be estimated by mean

$$\psi = \frac{1}{f} \sum_{i=1}^{f} \psi_i \tag{39}$$

According to the motion model, the f+1st two-view rotation axis vector n_{f+1} is $R(l, (f-i+1)\phi)n_i$ for any $1 \leqslant i \leqslant f$. In the presence of noise, we use the weighted sum over all previous two-view motions to predict the next two-view rotation axis vector:

$$n_{f+1}=\frac{1}{\sum_{i=1}^{f} i^{-1}}\sum_{i=1}^{f}(f-i+1)^{-1}R(l,(f-i+1)\phi)n_i \tag{40}$$

From (11) the next position of point X_f in frame F_f is predicted by

$$X_{f+1}=R(n_{f+1},\psi)(X_f-Q_f)+Q_{f+1} \tag{41}$$

where Q_f and Q_{f+1} are determined by (7). The prediction can be made for more than $p \geqslant 2$ frames by using the following equations recursively

$$n_{f+p}=R(l,\phi)n_{f+p-1} \tag{42}$$

$$X_{f+p}=R(n_{f+p},\psi)(X_{f+p-1}-Q_{f+p-1})+Q_{f+p} \tag{43}$$

The variances of samples in the summations of (37), (38) and (39) as well as the residuals in (35) and (14) indicate the accuracy of the model for the current set of frames. They also depend on the noise level.

If the object was occluded in parts of image sequences, the positions and the orientations of the object as well as the locations of the feature points on the object can be recovered by interpolation similar to the prediction procedure discussed above. For the motion of the rotation center, occlusion just means that some rows in the coefficient equations are missing. The solution can still be found if we have enough rows. For the precession part of the motion, the interpolation can be made in a way similar to prediction or extrapolation. When making interpolation we use both the "history" and the "future" of the missing part. For prediction, only the "history" is available. Furthermore, we can also extrapolate backwards to find "history", *i.e.*, to recall what has not been seen before. The essential assumption is that the motion is smooth.

4. MONOCULAR VISION

For the monocular case, 3-D positions can be predicted only up to a scale factor. However the image coordinates of a point can be predicted without knowing the scale factor.

For simplicity, we assume the model covers four frames. Let (X_i, Y_i) be the image coordinates of a 3-D point $P_i = (x_i, y_i, z_i)^t$, at time instant i; $i = 0, 1, 2, 3$. Without loss of generality, assume that the focal length of the camera is unity. The image coordinates and 3-D coordinates are related by

$$X_i = \frac{x_i}{z_i} \qquad Y_i = \frac{y_i}{z_i} \tag{44}$$

Let P_{i-1} and P_i be the corresponding points at time t_{i-1} and t_i, respectively. From (11) we have,

$$P_i = R_i P_{i-1} + T_i \tag{45}$$

where $R_i = (r_j^{(i)})$, $T_i = (\Delta x_i, \Delta y_i, \Delta z_i)^t$. Let $\hat{T}_i = (\Delta \hat{x}_i, \Delta \hat{y}_i, \Delta \hat{z}_i)^t = T_i / \| T_i \|$. In the monocular case, R_i and $\hat{T}_i$ can be determined from image frames F_{i-1} and F_i. The range z_{i-1} is given by [17]:

$$\begin{aligned} z_{i-1} &= \frac{\Delta x_i - X_i \Delta z_i}{X_i (r_{31}^{(i)} X_{i-1} + r_{32}^{(i)} Y_{i-1} + r_{33}^{(i)}) - (r_{11}^{(i)} X_{i-1} + r_{12}^{(i)} Y_{i-1} + r_{13}^{(i)})} \\ &= \| T_i \| \frac{\Delta \hat{x}_i - X_i \Delta \hat{z}_i}{X_i (r_{31}^{(i)} X_{i-1} + r_{32}^{(i)} Y_{i-1} + r_{33}^{(i)}) - (r_{11}^{(i)} X_{i-1} + r_{12}^{(i)} Y_{i-1} + r_{13}^{(i)})} \\ &= \| T_i \| A_i \end{aligned} \tag{46}$$

where A_i are function of X_{i-1}, X_i, Y_{i-1}, R_i and $\hat{T}_i$.

$$\begin{aligned} z_i &= r_{31}^{(i)} x_{i-1} + r_{32}^{(i)} y_{i-1} + r_{33}^{(i)} z_{i-1} + \Delta z_i \\ &= z_{i-1} (r_{31}^{(i)} X_{i-1} + r_{32}^{(i)} Y_{i-1} + r_{33}^{(i)} + \Delta \hat{z}_i / A_i) \\ &= z_{i-3} \prod_{j=i-2}^{i} (r_{31}^{(j)} X_{j-1} + r_{32}^{(j)} Y_{j-1} + r_{33}^{(j)} + \Delta \hat{z}_j / A_i) \\ &= z_{i-3} G_i \end{aligned} \tag{47}$$

where G_i is function of $X_{i-3} \cdots X_i$, $Y_{i-3} \cdots Y_i$, $R_{i-2} \cdots R_i$, $\hat{T}_{i-2} \cdots \hat{T}_i$. It can be determined from F_{i-3} to F_i. So

$$x_i = z_i X_i = z_{i-3} G_i X_i$$
$$y_i = z_i Y_i = z_{i-3} G_i Y_i \qquad (48)$$
$$z_i = z_{i-3} G_i$$

From equations (48) it is clear that the 3-D coordinates of a point in image frame F_i can be determined up to a scale factor z_{i-3}. This is also true for points in $F_{i-3} \cdots F_{i-1}$. From (45), all translation vector T_k can be determined up to a scale factor z_{i-3}. From (13) all coefficient vectors can be determined up to the same scale factor z_{i-3}. So the rotation center can be predicted up to the same scale factor. There is no ambiguity about predicted rotation. Finally, from (44), in the predicted image coordinates, this unknown scale factor is canceled out. In other words, we can assume z_{i-3} to be any arbitrary non-zero number, in particular, 1, to predict the image coordinates of points.

One point should be mentioned here. In the binocular case the set of corresponding points may be different for different consecutive image frame pairs. This means that the point set used in point correspondences in some images are allowed to be invisible in other images. The same is still true for the monocular case. In (47), we are tracing the same point all the way back to F_{i-3}. This is only for simplicity. If we choose a point P_c whose range is c. Applying this point to (46) and letting z_{i-1} to be c, Equations (47) mean the length of translation $\| T_i \|$ is proportional to c. Using image coordinates to rewrite (45), we have

$$\begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix} z_i - R_i \begin{bmatrix} X_{i-1} \\ Y_{i-1} \\ 1 \end{bmatrix} z_{i-1} = \| T_i \| \hat{T}_i$$

The above equation means the ranges (z_i and z_{i-1}) of other points are proportional to the length of translation vector $\| T_i \|$ which is proportional to c. So, the range of the other points can communicate the constant c to other image frames to determine the ranges of the points in those frames as z_{i-1} does in equations (47).

5. SIMULATIONS

Two simulation experiments to test the analysis were performed on a VAX 11/780. In the first case the image data were computer generated. In the second case, binocular image sequences of a model airplane undergoing a smooth motion were recorded using video cameras.

In the first experiment, a sequence of 3-D coordinates of points on a moving cube were generated by a program. The motion of the rotation center is characterized by coefficient vectors a_i in (7). By Theorem 3, the rotation of the body is about the two-view rotation axis n by angle θ between consecutive frames. The two-view rotation axis rotates about a fixed precession vector l by angle ϕ. The object is assumed to be a transparent cube of side length 10cm. Simulated cameras are 100cm away from the object. The viewing lines from the two cameras to the object form an angle of 45°. The cube covered about half of the image. The feature points used for point correspondences are the vertices of the cube. As the object is undergoing motion, these points generate a sequence of 3-D coordinates. These 3-D coordinates are digitized by simulated cameras of resolution from 64 by 64 up to 512 by 512. Because of digitization errors, the set of points no longer satisfy the rigid body constraint. To find the best two-view motion parameters, a least squares solution is obtained [7].

In the experiments presented, the following motion parameters are used. Precession vector: $l=(0, 0, 1)^t$; Precession angular frequency: $\phi=0.4$; Two-view rotation axis of first two-view motion: $n_1=\frac{1}{\sqrt{17}}(1, 0, 4)^t$, Coefficient vectors: $a_1=(-2, -3, -1)^t$, $a_2=(0.5, 0.5, 0.25)^t$, $a_3=(5^{-3}, 5^{-3}, 2.5^{-3})^t$. The results given are the mean values from 20 trials, each of which randomly chooses the original orientation of the cube.

Fig.7 describes the estimation errors of l and ϕ. The errors of l are defined as the length of the difference of the estimated and the real unit vectors. The errors approach zero very fast as the number of two-view motions f increases. Based on the estimated model parameters, the predictions of the 3-D coordinates of each point at the next several time instants are made. The predicted

and the actual digitized positions are compared. Their distances are considered as prediction errors. The relative errors are defined as the prediction errors divided by the length of the diagonal of the cube. The results of predicting more frames are presented in Fig.8. Fig.9 shows the mean relative prediction errors at time t_5 as a function of image resolution. Fig.9 indicates also the errors as the function of the number of point correspondences. Fig.10 shows the mean relative prediction errors as function of image resolution and the number of two-view motions covered. Fig.11 gives the mean relative prediction errors as the function of the number of point correspondences and the number of two-view motions covered.

Data for the second experiments were taken from a model airplane. The setup of the real cameras is the same as that in the simulation experiment discussed above. Without translation, the model airplane rotates about a vertical line by about 15° per frame. Simultaneously it also rotates about its head-tail central line by about 8° per frame as shown in Fig.12. The feature points used for point correspondences are at the tip of the wings. The feature points and their correspondences are determined manually. The number of point correspondences is 4. Starting from the fourth frame, the prediction is made for the 3-D coordinates of the feature points at the next time instant. The relative maximum prediction errors are shown in Fig.12. The relative errors are defined by the error divided by the maximum distance between feature points. The camera resolution is 512 by 512. Comparing with the synthetic data case, the additional error source is lens distortion. The relative errors in this experiment are close to the synthetic data case.

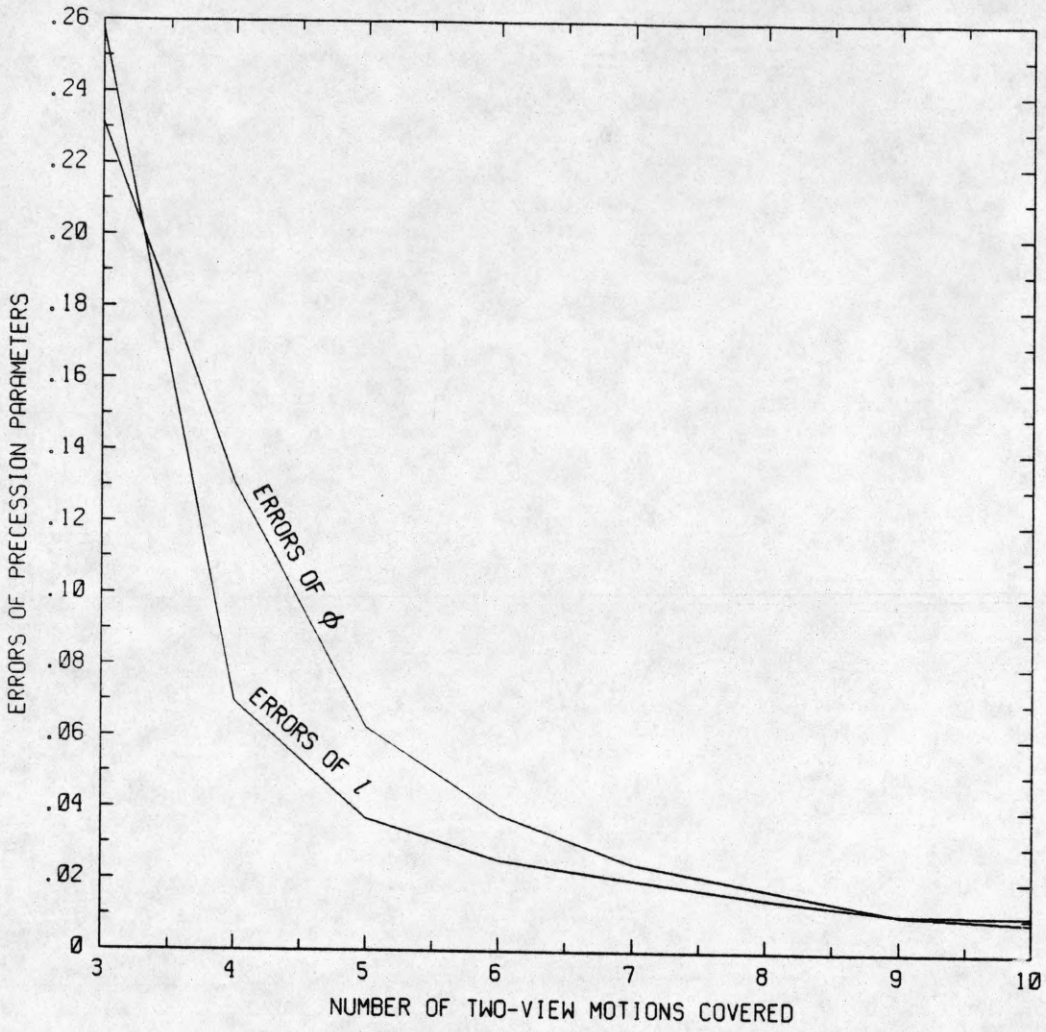

Fig.7 Errors of l (precession vector) and ϕ (precession angular angular frequency) as function of number of two-view motions covered

number of point correspondences : 3
frame predicted : next
image resolution : 256 by 256

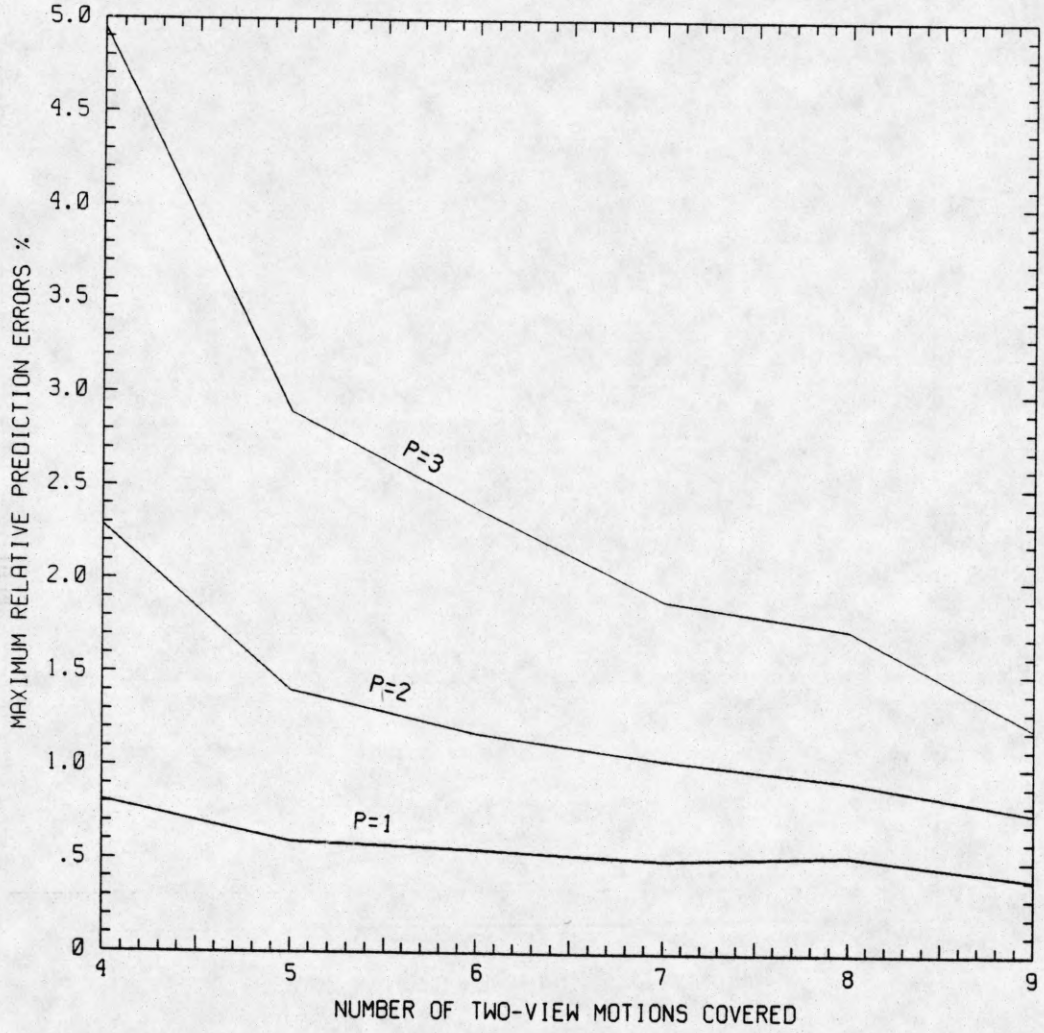

Fig.8 Maximum relative prediction errors (%) as function of number of two-view motions covered and number of frames predicted (p).

frame predicted : next p-th frame
number of point correspondences : 3
image resolution : 512 by 512
degree of polynomial k-1 in equation (7): 2

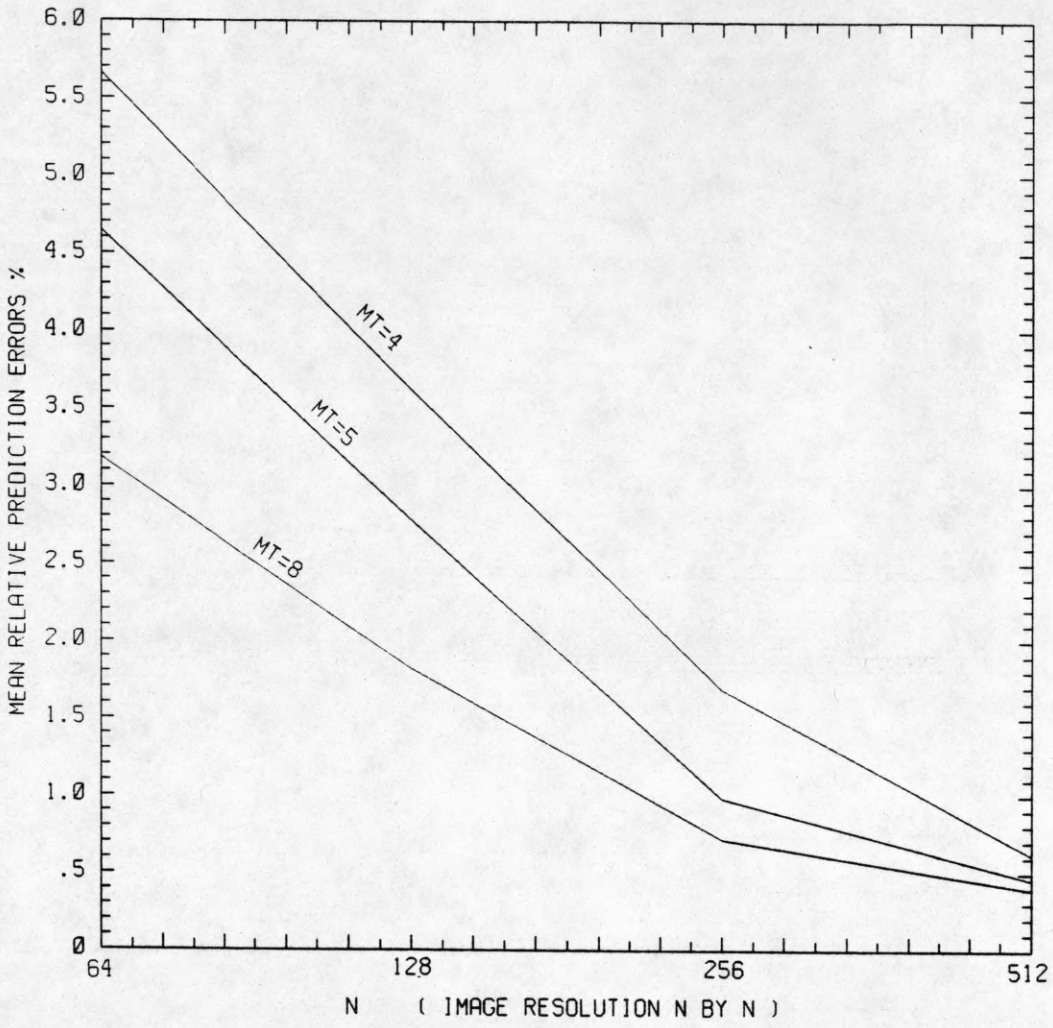

Fig.9 Mean relative prediction errors (%) as function of
image resolution (n by n) and number of point correspondences (npts)

frame predicted: next frame
number of two-view motions covered: 5
degree of polynomial k-1 in equation (7): 2

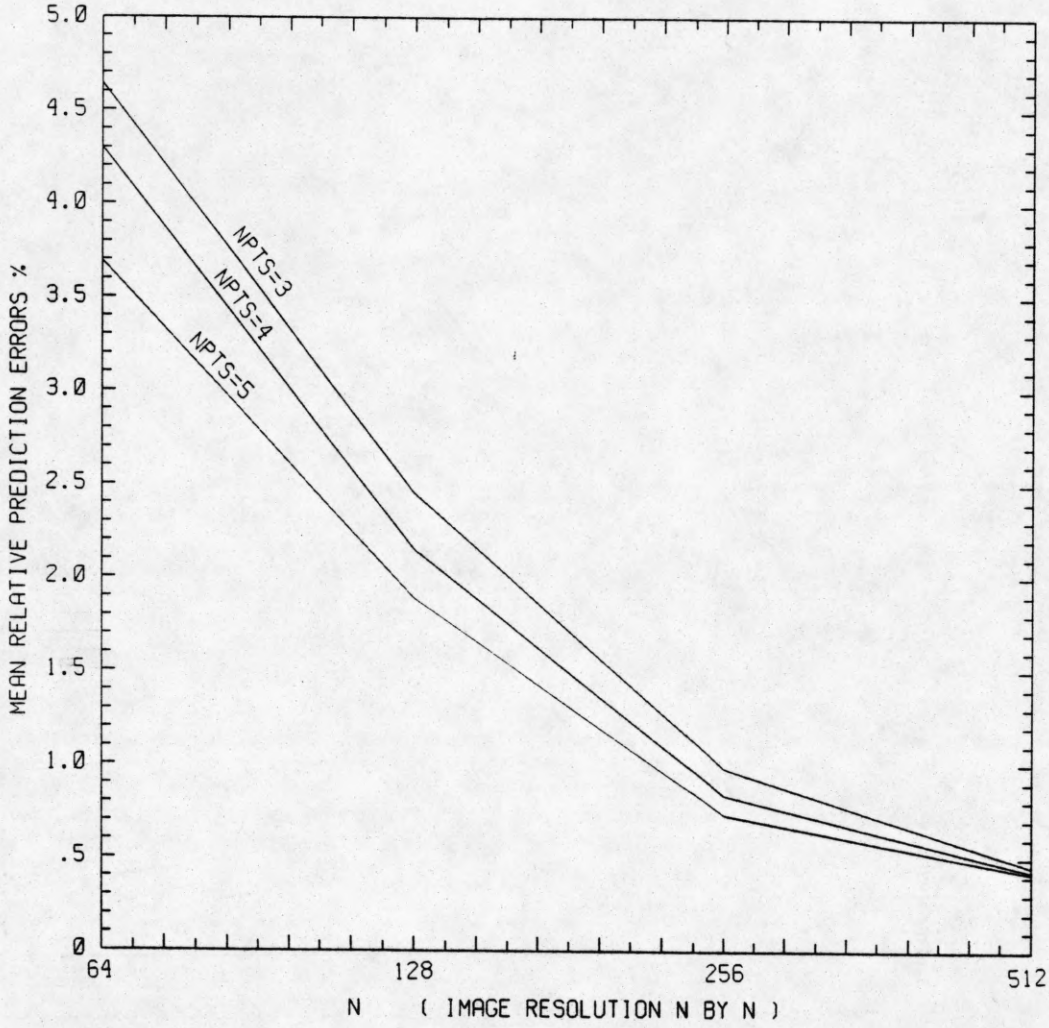

Fig.10 Mean relative prediction errors (%) as function of
image resolution (n by n) and number of two-view motions covered (mt)

frame predicted: next frame
number of point correspondences: 3
degree of polynomial k-1 in equation (7): 2

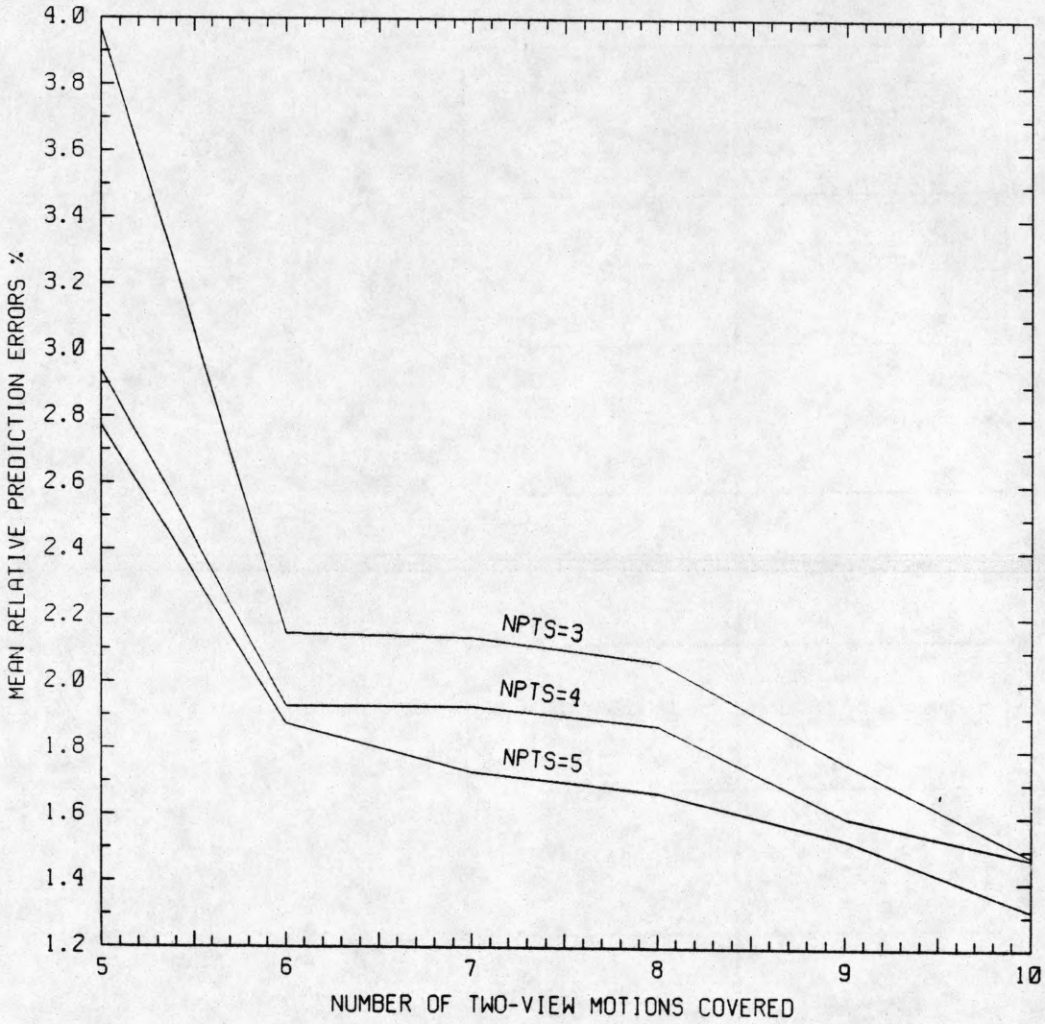

Fig.11 Mean relative prediction errors (%) as function of
number of point correspondences (npts) and number of two-view motions covered

frame predicted: next frame
image resolution: 128 by 128
degree of polynomial k-1 in equation (7): 3

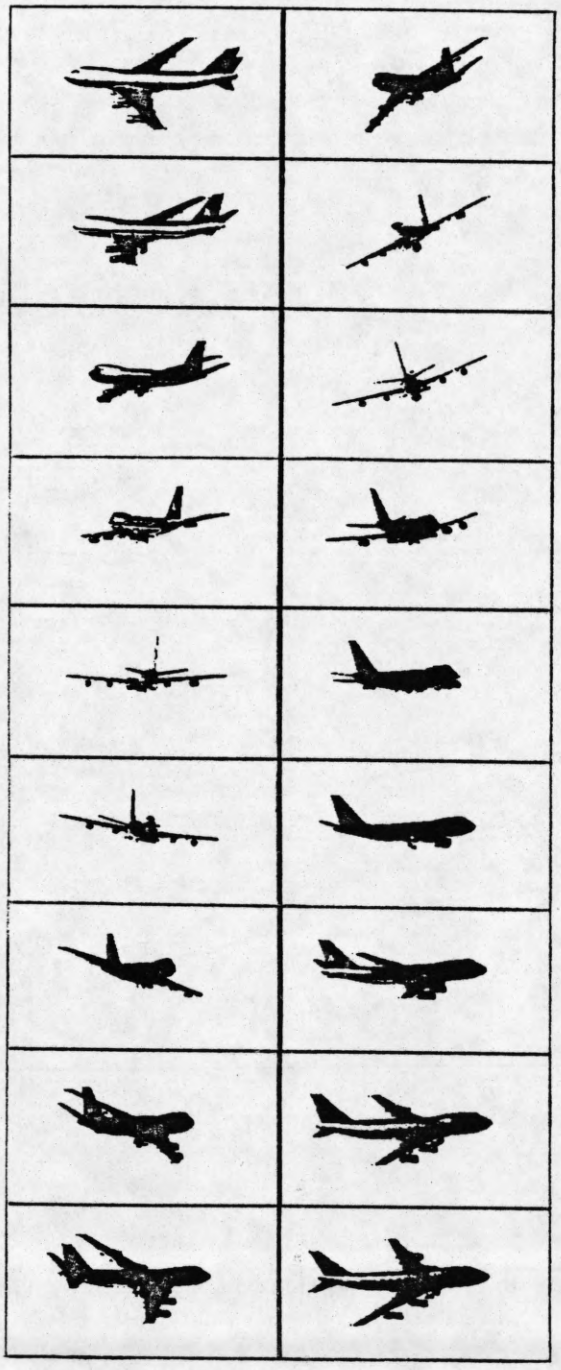

Fig.12 Image Sequence of a model airplane

left column: left view; right column: right view. (From the fifth frame, the maximum relative prediction errors are shown to the left of the corresponding frame pairs.)

6. SUMMARY

We have described an approach to modeling and estimating general 3-D motion of an object from image sequences. The dynamics of the moving object is modeled by two components. First, the rotation of the object is assumed to be a precession which can be modeled by such motion: the rotation about a fixed-on-body axis that rotates about a spatially fixed axis. One of the conditions of object dynamics leading to such motion is that the object under motion is symmetric and its angular momentum is constant. Second, the object is assumed to undergo a smooth translational motion. In particular, we assume that the location of the rotation center of the object can be represented by a vector polynomial in time. The motion of any points on the object can be expressed by the superposition of these two components. The problem of modeling motion then amounts to estimating the parameters of precession with respect to the rotation center, and the parameters of translation of the rotation center. This estimation can be performed from either point correspondences or optical flow over a sequence of image frames. We use the former to discuss. Using the technique of two-view motion analysis, estimates of two-view rotational and translational parameters can be derived. To reduce the sensitivity to noise, least squares estimates are obtained from multiple features in two-view motion analysis. Based on the parameters of two-view motion, the parameters of LCAM model are estimated so as to understand the local motion and predict the future motion. Again, least squares method is used for model parameter estimation from multiple image frames to combat noise.

We have presented a linear algorithm that implements our approach. The experiments have been performed on image sequences obtained from simulated as well as actual moving objects. To test the accuracy of the model, the predictions of the locations of object points were obtained and the errors between the predicted and actual locations were measured. The errors of estimated model parameters have been presented for different numbers of image frames. The prediction errors have been shown for different image resolutions, different numbers of object points, different numbers of image frames covered, and different numbers of frames predicted.

APPENDICES

APPENDIX 1.

THEOREM 1. Let f=k. Define S_k^j to be a 3 by 3 matrix :

$$S_k^j=\sum_{l=0}^{k-j}(-1)^l\binom{k}{l}R_{k-l}R_{k-l-1}\cdots R_{j+1}I \qquad j=0,1,2,\cdots,k$$

Define number u_{ij} :

$$u_{ij}=\sum_{m=1}^{i}(-1)^{i-m}m^{j-1}\binom{i}{m} \qquad j=i+1,i+2,\cdots,k$$

Then

$$S_k^0a_1=-\sum_{l=1}^{k}S_k^lT_l$$

$$a_k=\frac{-1}{(k-1)!}\left(\sum_{m=1}^{k-1}S_{k-1}^mT_m+S_{k-1}^0a_1\right)$$

$$a_{k-1}=\frac{-1}{(k-2)!}\left(\sum_{m=1}^{k-2}S_{k-2}^mT_m+S_{k-2}^0a_1+u_{k-2\,k}a_k\right)$$

$$a_{k-2}=\frac{-1}{(k-3)!}\left(\sum_{m=1}^{k-3}S_{k-3}^mT_m+S_{k-3}^0a_1+u_{k-3\,k}a_k+u_{k-3\,k-1}a_{k-1}\right)$$

.....

$$a_3=\frac{-1}{2!}\left(\sum_{m=1}^{2}S_2^mT_m+S_2^0a_1+u_{2\,k}a_k+u_{2\,k-1}a_{k-1}+\cdots+u_{2\,4}a_4\right)$$

$$a_2=\frac{-1}{1!}\left(\sum_{m=1}^{1}S_1^1T_m+S_1^0a_1+u_{1\,k}a_k+u_{1\,k-1}a_{k-1}+\cdots+u_{1\,3}a_3\right)$$

To prove Theorem 1, we first prove some lemmas.

LEMMA 1. Define

$$\hat{S}_k^j=\sum_{l=0}^{k-j}(-1)^l\binom{k}{l}R_{k-l+1}\cdots R_{j+2}I \qquad \hat{S}_k^k=I$$

Then,

$$\hat{S}_k^{\,0} R_1 - S_k^{\,0} = S_k^{\,0}{}_{+1}$$

Proof.

$$\hat{S}_k^{\,0} R_1 - S_k^{\,0} = \sum_{l=0}^{k} (-1)^l \binom{k}{l} R_{k-l+1} \cdots R_2 R_1 I - \sum_{l=0}^{k} (-1)^l \binom{k}{l} R_{k-l} \cdots R_1 I$$

$$= \binom{k}{0} R_{k+1} \cdots R_2 R_1 I + \sum_{l=1}^{k} (-1)^l \binom{k}{l} R_{k-l+1} \cdots R_2 R_1 I - \sum_{l=0}^{k-1} (-1)^l \binom{k}{l} R_{k-l} \cdots R_1 I - (-1)^k \binom{k}{k} I$$

$$= \binom{k+1}{0} R_{k+1} \cdots R_2 R_1 I + \sum_{l=1}^{k} (-1)^l \binom{k}{l} R_{k-l+1} \cdots R_2 R_1 I$$

$$+ \sum_{m=1}^{k} (-1)^m \binom{k}{m-1} R_{k-m+1} \cdots R_1 I + (-1)^{k+1} \binom{k+1}{k+1} I \qquad \text{(A.1)}$$

Using $\binom{k}{l} + \binom{k}{l-1} = \binom{k+1}{l}$, (A.1) becomes

$$\binom{k+1}{0} R_{k+1} \cdots R_2 R_1 I + \sum_{l=1}^{k} (-1)^l \binom{k+1}{l} R_{k-l+1} \cdots R_2 R_1 I + (-1)^{k+1} \binom{k+1}{k+1} I$$

$$= \sum_{l=0}^{k+1} (-1)^l \binom{k+1}{l} R_{k-l+1} \cdots R_2 R_1 I$$

$$= S_k^{\,0}{}_{+1}$$

□

LEMMA 2.

$$\hat{S}_k^{\,j-1} - S_k^{\,j} = S_k^{\,j}{}_{+1}$$

Proof.

$$\hat{S}_k^{\,j-1} - S_k^{\,j} = S_k^{\,j}{}_{-1} = \sum_{l=0}^{k-j+1} (-1)^l \binom{k}{l} R_{k-l+1} \cdots R_{j+1} I - \sum_{l=0}^{k-j} (-1)^l \binom{k}{l} R_{k-l} \cdots R_{j+1} I$$

$$= \binom{k}{0} R_{k+1} \cdots R_{j+1} I + \sum_{l=1}^{k-j+1} (-1)^l \binom{k}{l} R_{k-l+1} \cdots R_{j+1} I + \sum_{m=1}^{k-j+1} (-1)^m \binom{k}{m-1} R_{k-m+1} \cdots R_{j+1} I \qquad \text{(A.2)}$$

Using $\binom{k}{l} + \binom{k}{l-1} = \binom{k+1}{l}$, (A.2) becomes

$$\binom{k+1}{0} R_{k+1} \cdots R_2 R_{j+1} I + \sum_{l=1}^{k+1-j} (-1)^{(k+1}_{l}) R_{k+1-l} \cdots R_{j+1} I$$

$$= \sum_{l=0}^{k+1-j} (-1)^l \binom{k+1}{l} R_{k-l+1} \cdots R_{j+1} I$$

$$= S_{k+1}^{j}$$

□

LEMMA 3. Let $k \geq 2$, $1 \leq m \leq k-1$. Then

$$\sum_{j=1}^{k} (-1)^{k-j} j^m \binom{k}{j} = 0$$

$$\sum_{j=1}^{k} (-1)^{k-j} j^k \binom{k}{j} = k!$$

Proof. From the binomial equation

$$(1+x)^k = \sum_{j=0}^{k} \binom{k}{j} x^j \tag{A.3}$$

By differentiating both sides, then multiplying by x, we have

$$(1+x)^{k-1} kx = \sum_{j=1}^{k} j \binom{k}{j} x^j$$

Differentiating both sides and then multiplying by x again, we get

$$(1+x)^{k-2}(k(k-1)x^2 + k(1+x)x) = \sum_{j=1}^{k} j^2 \binom{k}{j} x^j$$

Generally, doing this m times, $1 \leq m \leq k$, we get

$$(1+x)^{k-m}[k(k-1)\cdots(k-m+1)x^m + (1+x)P(x)] = \sum_{j=1}^{k} j^m \binom{k}{j} x^j \tag{A.4}$$

where $P(x)$ is an $(m-1)$-th degree polynomial. This can be readily proved by induction on m. Let $x = -1$. For $1 \leq m \leq k-1$, (A.4) gives

$$0 = \sum_{j=1}^{k} j^m \binom{k}{j} (-1)^j = \sum_{j=1}^{k} (-1)^{k-j} j^m \binom{k}{j}$$

Let $m = k$, (A.4) gives

$$k!(-1)^k = \sum_{j=1}^{k} (-1)^j \, j^m \binom{k}{j}$$

Multiplying both sides by $(-1)^k$. Lemma 3 is proved. □

LEMMA 4. From equation (14), let $D = [d_{ij}]$.

$$d_{ij} = i^{j-1} I - (i-1)^{j-1} R_i$$

Define

$$\hat{d}_{ij} = i^{j-1} I - (i-1)^{j-1} R_{i+1}$$

Then

$$\sum_{j=1}^{k} S_k^j d_{j\,k+1} = k! I$$

$$\sum_{j=1}^{k} \hat{S}_k^j \hat{d}_{j\,k+1} = k! I$$

Proof. We need only prove the first equation. The difference between the first equation and the second one is that the subscripts of rotation matrices R_i are all incremented by 1 in the second equation.

$$\sum_{j=1}^{k} S_k^j d_{j\,k+1}$$

$$= \sum_{j=1}^{k} S_k^j (j^k I - (j-1)^k R_j)$$

$$= \sum_{j=1}^{k} j^k S_k^j - \sum_{j=2}^{k} (j-1)^k S_k^j R_j$$

$$= \sum_{j=1}^{k} j^k S_k^j - \sum_{m=1}^{k-1} m^k S_k^{m+1} R_{m+1}$$

$$= \sum_{j=1}^{k} j^k S_k^j - \sum_{j=1}^{k-1} j^k \sum_{l=0}^{k-j-1} (-1)^l \binom{k}{l} R_{k-l} \cdots R_{j+2} R_{j+1} I$$

$$= \sum_{j=1}^{k} j^k S_k^j - \sum_{j=1}^{k-1} j^k \left[\sum_{l=0}^{k-j} (-1)^l \binom{k}{l} R_{k-l} \cdots R_{j+2} R_{j+1} I - (-1)^{k-j} \binom{k}{j} I \right]$$

$$= k^k I + \sum_{j=1}^{k-1} j^k S_k^j - \sum_{j=1}^{k-1} j^k S_k^j + \sum_{j=1}^{k-1} (-1)^{k-j} j^k \binom{k}{j} I$$

$$=(-1)^{k-k}k^k\binom{k}{k}+\sum_{j=1}^{k-1}(-1)^{k-j}j^k\binom{k}{j}I$$

$$=\sum_{j=1}^{k}(-1)^{k-j}j^k\binom{k}{j}I$$

By Lemma 3 the above equation is equal to $k!I$

□

Now we are ready to prove Theorem 1.

Proof of Theorem 1. We first prove

$$S_k^0 a_1=-\sum_{j=1}^{k}S_k^j T_j \tag{A.5}$$

by induction on k.

Let $k=1$.

$$S_1^0=R_1-I$$

$$-\sum_{j=1}^{1}S_1^j T_j=-S_1^1 T_1=-IT_1=-T_1$$

Equation

$$S_1^0 a_1=-\sum_{j=1}^{1}S_1^j T_j$$

means $(R_1-I)a_1=-T_1$. This is $DA=T$ when $k=1$. So, the equation (A.5) is true when $k=1$.

Assume (A.5) is true up to k. Let D be $k+1$ by $k+1$ matrix (whose elements are 3 by 3 matrices). From first k equations of $DA=T$, we have

$$d_{11}a_1+d_{12}a_2+\cdots+d_{1k}a_k=T_1-d_{1\,k+1}a_{k+1}$$

$$d_{21}a_1+d_{22}a_2+\cdots+d_{2k}a_k=T_2-d_{2\,k+1}a_{k+1}$$

.....

$$d_{k1}a_1+d_{k2}a_2+\cdots+d_{kk}a_k=T_k-d_{k\,k+1}a_{k+1}$$

By induction hypothesis, we have

$$S_k^0 a_1 = -\sum_{j=1}^{k} S_k^j (T_j - d_{j\,k+1} a_{k+1}) \tag{A.6}$$

From (7), let

$$Q(i) = a_1 + a_2 i + \cdots + a_k i^{k-1} + a_{k+1} i^k \tag{A.7}$$

Consider another polynomial $\hat{P}(i)$ with t_1 as initial starting time. *i.e.*

$$\hat{P}(i) = P(i+1) \tag{A.8}$$

$$\hat{Q}(i) = \hat{a}_1 + \hat{a}_2 i + \cdots + \hat{a}_k i^{k-1} + \hat{a}_{k+1} i^k \tag{A.9}$$

We have

$$\hat{a}_1 + \hat{a}_2 i + \hat{a}_3 i^2 + \cdots + \hat{a}_{k+1} i^k$$

$$= a_1 + a_2 (i+1) + a_3 (i+1)^2 + \cdots + a_{k+1} (i+1)^k$$

From (A.7), (A.8) and (A.9), the coefficient of variable i should be equal. We have

$$\hat{a}_1 = a_1 + a_2 + \cdots + a_{k+1}, \qquad \hat{a}_{k+1} = a_{k+1} \tag{A.10}$$

For the Polynomial in (A.9), we get the coefficient equations :

$$\hat{D}\hat{A} = \hat{T}$$

where

$$\hat{D} = [\hat{d}_{ij}]_{k \times (k+1)}$$

$$\hat{d}_{ij} = i^{j-1} I - (i-1)^{j-1} R_{i+1} \qquad 1 \leqslant i \leqslant k \quad 1 \leqslant j \leqslant k+1$$

$$\hat{A} = (\hat{a}_1, \cdots, \hat{a}_k)^t$$

$$\hat{T} = (T_2, \cdots, T_{k+1})^t$$

The equation $\hat{D}\hat{A} = \hat{T}$ can be derived directly from $DA = T$. The difference of $\hat{D}\hat{A} = \hat{T}$ from $DA = T$ is that the unknowns have hat, and other subscriptions are incremented. Using the induction hypothesis on $\hat{D}\hat{A} = \hat{T}$, we have

$$\hat{S}_k^{\,0}\hat{a}_1=-\sum_{j=1}^{k}\hat{S}_k^{j}(T_{j+1}-\hat{d}_{j+1\,k+1}\hat{a}_{k+1}) \tag{A.11}$$

From (A.10) and the first equations in (13), we get

$$\hat{a}_1=a_1+a_2+\cdots+a_{k+1}=R_1a_1+T_1$$

So,

$$R_1a_1=\hat{a}_1-T_1 \tag{A.12}$$

We get

$$\hat{S}_k^{\,0}R_1a_1=\hat{S}_k^{\,0}(\hat{a}_1-T_1)=\hat{S}_k^{\,0}\hat{a}_1-\hat{S}_k^{\,0}T_1 \tag{A.13}$$

From (A.10) and (A.11), (A.13) becomes

$$\hat{S}_k^{\,0}R_1a_1=-\sum_{j=1}^{k}\hat{S}_k^{j}T_{j+1}+\sum_{j=1}^{k}\hat{S}_k^{j}\hat{d}_{j\,k+1}a_{k+1}-\hat{S}_k^{\,0}T_1$$

$$=-\sum_{j=0}^{k}\hat{S}_k^{j}T_{j+1}+\sum_{j=1}^{k}\hat{S}_k^{j}\hat{d}_{j\,k+1}a_{k+1}$$

From Lemma 4, we have

$$\hat{S}_k^{\,0}R_1a_1=-\sum_{j=0}^{k}\hat{S}_k^{j}T_{j+1}+k!a_{k+1} \tag{A.14}$$

From (A.6), we get

$$S_k^{\,0}a_1=-\sum_{j=1}^{k}S_k^{j}T_j+\sum_{j=1}^{k}S_k^{j}d_{j\,k+1}a_{k+1}$$

By Lemma 4, we have

$$S_k^{\,0}a_1=-\sum_{j=1}^{k}S_k^{j}T_j+k!\,a_{k+1} \tag{A.15}$$

Using Lemma 1, (A.14) and (A.15), we get

$$S_{k+1}^{\,0}a_1=(\hat{S}_k^{\,0}R_1-S_k^{\,0})a_1$$

$$=\hat{S}_k^{\,0}R_1a_1-S_k^{\,0}a_1$$

$$=-\sum_{j=0}^{k}\hat{S}_k^j T_{j+1}+\sum_{j=1}^{k}S_k^j T_j$$

$$=-\sum_{l=1}^{k+1}\hat{S}_k^{l-1}T_l+\sum_{j=1}^{k}S_k^j T_j$$

$$=-\sum_{j=1}^{k}(\hat{S}_k^{j-1}-S_k^j)T_j-T_{k+1}$$

$$=-\sum_{j=1}^{k+1}S_k^{j}{}_{+1}T_j \qquad \text{(A.16)}$$

So, equation (A.5) is true for $k+1$. Thus, (A.5) is proved for all $k \geqslant 1$.

Then let us prove the rest of the equations. Let $2 \leqslant i \leqslant k-1$. Considering the first i equations in (13), we move $a_{i+1}, \cdots, a_k$ to the right side. Using (A.5) to solving a_1, we get

$$S_i^0 a_1=-\sum_{m=1}^{i}S_i^m(T_m-\sum_{j=i+1}^{k}d_{mj}a_j)=-\sum_{m=1}^{i}S_i^m T_m-\sum_{j=i+1}^{k}(\sum_{m=1}^{i}S_i^m d_{mj})a_j \qquad \text{(A.17)}$$

Evaluate the term

$$\sum_{m=1}^{i}S_i^m d_{mj}=\sum_{m=1}^{i}S_i^m(m^{i-1}I-(m-1)^{j-1}R_m)$$

$$=\sum_{m=1}^{i-1}m^{j-1}S_i^m-\sum_{m=2}^{i}(m-1)^{j-1}R_m+i^{j-1}I$$

$$=\sum_{m=1}^{i-1}m^{j-1}\sum_{l=0}^{i-m}(-1)^l\binom{i}{l}R_{i-l}\cdots R_{m+1}I-\sum_{m=2}^{i}(m-1)^{j-1}\sum_{l=0}^{i-m}(-1)^l\binom{i}{l}R_{i-l}\cdots R_{m+1}R_m I+i^{j-1}I$$

$$=\sum_{m=1}^{i-1}m^{j-1}\sum_{l=0}^{i-m-1}(-1)^l\binom{i}{l}R_{i-l}\cdots R_{m+1}I-\sum_{v=2}^{i-1}(v-1)^{j-1}\sum_{l=0}^{i-v-1}(-1)^l\binom{i}{l}R_{i-l}\cdots R_{v+1}R_v I$$

$$+\sum_{m=1}^{i-1}(-1)^{i-m}m^{j-1}\binom{i}{i-m}I+i^{j-1}I$$

$$=[(\sum_{m=1}^{i-1}(-1)^{i-m}m^{j-1}\binom{i}{m}+i^{j-1})]I$$

$$=[\sum_{m=1}^{i}(-1)^{i-m}m^{j-1}\binom{k}{j}]I$$

$$=u_{ij}I$$

So, (A.17) becomes

$$S_i^{\,0} a_1 = -\sum_{m=1}^{i} S_i^m T_m - \sum_{j=i+1}^{k} u_{ij}\, a_j$$

Or,

$$u_{i\,i+1} a_{i+1} = -\left(\sum_{m=1}^{i} S_i^m T_m + S_i^{\,0} a_1 + \sum_{j=i+2}^{k} u_{ij}\, a_j \right)$$

By Lemma 3, $u_{i\,i+1} = i!$. So we have

$$a_{i+1} = \frac{-1}{i!}\left(\sum_{m=1}^{i} S_i^m T_m + S_i^{\,0} a_1 + \sum_{j=i+2}^{k} u_{ij}\, a_j \right) \qquad \text{(A.18)}$$

Letting $i = k-1, k-2, \cdots, 2, 1$, we get the equations in Theorem 1.

□

APPENDIX 2

THEOREM 2. In the case of rotation without precession, let w be any vector parallel to the rotation axes. Then,

$$S_k^0 w = 0 \tag{A.19}$$

and for any vector a

$$(S_k^0 a) \cdot w = 0 \tag{A.20}$$

Proof. We prove (A.20). The proof of equation (A.19) is very similar. We are to prove another equation at the same time: Under the same condition, for any vector b the following equation holds

$$(\hat{S}_k^0 b) \cdot w = 0 \tag{A.21}$$

We use induction on k.

Let $k=1$. $(S_1^0 a) \cdot w = ((I - R_1)a) \cdot w = (a - R_1 a) \cdot w$. R_1 is rotation matrix with rotation axis w. So the difference of the original vector a and the rotated vector $R_1 a$ is orthogonal to the rotation axis. *i.e.*, $(a - R_1 a) \cdot w = 0$. Similarly $(\hat{S}_1^0 b) \cdot w = 0$.

Assuming $(S_k^0 a) \cdot w = 0$ $(\hat{S}_k^0 b) \cdot w = 0$ for $k \geqslant n$. Let $k = n+1$. From Lemma 1, $S_{n+1}^0 a = \hat{S}_n^0 R_1 - S_n^0$. We have

$$(S_{n+1}^0 a) \cdot w = (\hat{S}_n^0 R_1 a - S_n^0 a) \cdot w = (\hat{S}_n^0 R_1 a) \cdot w - (S_n^0 a) \cdot w = 0 + 0 = 0$$

Similarly (A.21) holds for $k = n+1$.

□

APPENDIX 3

Let R_1 and R_2 be rotation matrices with rotation axes n_1 and n_2, rotation angle θ_1 and θ_2, respectively. $R_2R_1-2R_1+I$ is nonsingular if and only if $n_1 \neq n_2$ and $\theta_1 \neq 0$, $\theta_2 \neq 0$.

Proof. Let $R_2R_1-2R_1+I=A$. A is nonsingular if and only if there exist three vectors U_0, V_0, W_0 such that AU_0, AV_0, AW_0 are independent. Notice that

$$AU_0=(R_2R_1U_0-R_1U_0)-(R_1U_0-U_0)$$

Let $R_1U_0=U_1$, $R_2R_1U_0=U_2$. i.e.: U_1 is a vector rotated by R_1 from U_0. U_2 is a vector rotated from U_1 by rotation matrix R_2. Let S_1 be the plane perpendicular to n_1, S_2 be the plane perpendicular to n_2. The two planes intersect each other at an angle ϕ as shown in Fig.13. Construct a Cartesian coordinate system as shown in Fig.13.

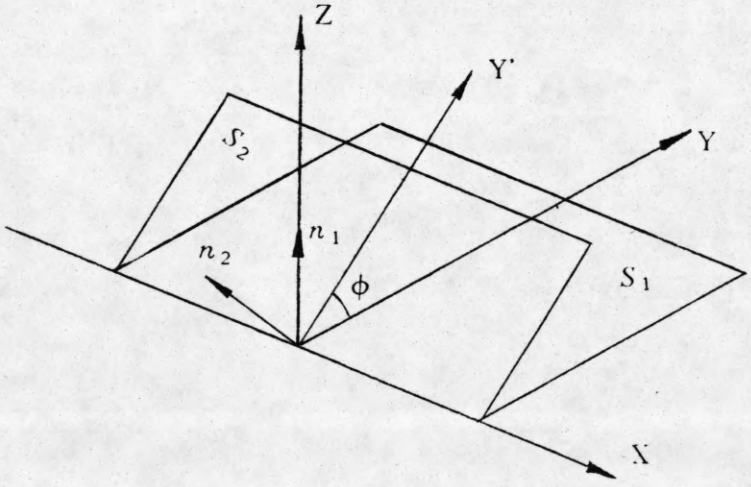

Fig.13 Two rotation planes and coordinate systems

The X-axis lies in the intersection line of S_1 with S_2. The Z-axis is in the same direction as n_1. S_1 then lies in X-Y plane. Let

$$R_1U_0=(1,0,0)^t, \quad R_1V_0=(0,1,0)^t, \quad R_1W_0=(0,0,1)^t$$

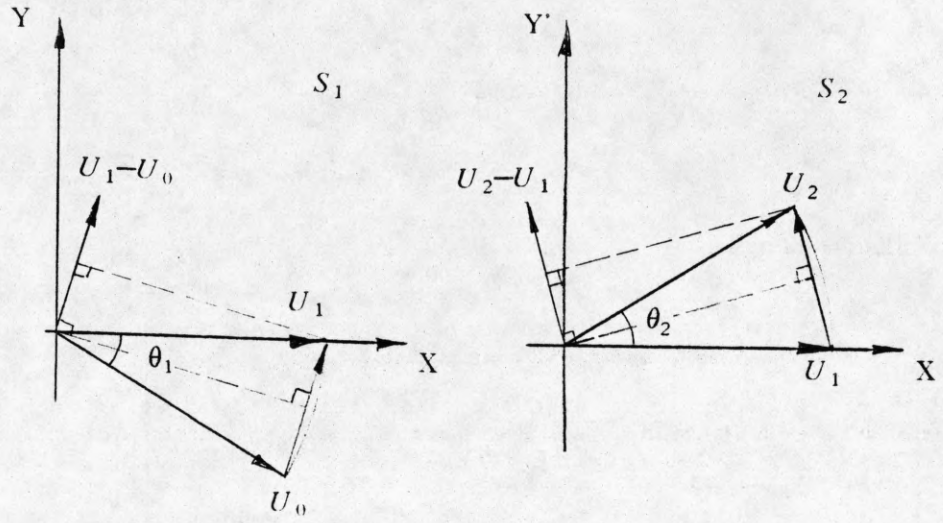

Fig.14 Vectors rotated from U_0

As shown in Fig.14, the vector U_1-U_0 lies in the X-Y plane. The angle from U_1-U_0 to the Y-axis is $\frac{\theta_1}{2}$. The length of U_1-U_0 is $2\sin\frac{\theta_1}{2}$. Similarly, in plane S_2, let Y' be the axis corresponding to projection of Y axis on to the plane S_2. $U_2-U_1=R_2U_1-U_1$ lies in S_2. The angle from Y' axis to U_2-U_1 is $\theta_2/2$. The length of U_2-U_1 is $2\sin\frac{\theta_2}{2}$. Let the italics X, Y, Y' be unit vectors on the coordinate axes. We have

$$U_1-U_0=2\sin^2\frac{\theta_1}{2}X+2\sin\frac{\theta_1}{2}\cos\frac{\theta_1}{2}Y \tag{A.22}$$

$$U_2-U_1=-2\sin^2\frac{\theta_2}{2}X+2\sin\frac{\theta_2}{2}\cos\frac{\theta_2}{2}Y' \tag{A.23}$$

Similarly we have

$$V_1-V_0=-2\sin\frac{\theta_1}{2}\cos\frac{\theta_1}{2}X+2\sin^2\frac{\theta_1}{2}Y \tag{A.24}$$

$$V_2-V_1=-2\cos\phi\,\sin\frac{\theta_2}{2}\cos\frac{\theta_2}{2}X-2\cos\phi\,\sin^2\frac{\theta_2}{2}Y' \tag{A.25}$$

$$AV_0=(V_2-V_1)-(V_1-V_0) \tag{A.26}$$

$$W_1-W_0=0 \tag{A.27}$$

$$WE_2-W_1=-2\sin\phi\,\sin\frac{\theta_2}{2}\cos\frac{\theta_2}{2}X-2\sin\phi\,\sin^2\frac{\theta_2}{2}Y' \tag{A.28}$$

$$AW_0=(W_2-W_1)-(W_1-W_0) \tag{A.29}$$

Using (A.22)-(A.29) we have

$$\begin{bmatrix}AU_0\\AV_0\\AW_0\end{bmatrix}=-2\begin{vmatrix}\sin^2\frac{\theta_1}{2}+\sin^2\frac{\theta_2}{2} & \sin\frac{\theta_1}{2}\cos\frac{\theta_1}{2} & -\sin\frac{\theta_2}{2}\cos\frac{\theta_2}{2}\\ \cos\phi\,\sin\frac{\theta_2}{2}\cos\frac{\theta_2}{2}-\sin\frac{\theta_1}{2}\cos\frac{\theta_1}{2} & \sin^2\frac{\theta_1}{2} & \cos\phi\,\sin^2\frac{\theta_2}{2}\\ \sin\phi\,\sin\frac{\theta_2}{2}\cos\frac{\theta_2}{2} & 0 & \sin\phi\,\sin^2\frac{\theta_2}{2}\end{vmatrix}\begin{bmatrix}X\\Y\\Y'\end{bmatrix}$$

$$=-2M\begin{bmatrix}X\\Y\\Y'\end{bmatrix}$$

$$|M|=\sin\phi\begin{vmatrix}\sin^2\frac{\theta_1}{2}+\sin^2\frac{\theta_2}{2} & \sin\frac{\theta_1}{2}\cos\frac{\theta_1}{2} & -\sin\frac{\theta_2}{2}\cos\frac{\theta_2}{2}\\ \cos\phi\,\sin\frac{\theta_2}{2}\cos\frac{\theta_2}{2}-\sin\frac{\theta_1}{2}\cos\frac{\theta_1}{2} & \sin^2\frac{\theta_1}{2} & \cos\phi\,\sin^2\frac{\theta_2}{2}\\ \sin\frac{\theta_2}{2}\cos\frac{\theta_2}{2} & 0 & \sin^2\frac{\theta_2}{2}\end{vmatrix}$$

Multiplying the third row by $\cos\phi$ and adding the result to the second row, we get

$$|M|=2\sin\phi\,\sin^2\frac{\theta_1}{2}\sin^2\frac{\theta_1}{2}$$

Assume $0\leqslant\theta_1<2\pi$, $0\leqslant\theta_2<2\pi$.

If M is singular, $\phi=0$ or $\theta_1=0$ or $\theta_2=0$.

If $\phi\neq0$, $\theta_1\neq0$, $\theta_2\neq0$, X, Y, and Y' are independent. So A is not singular.

ACKNOWLEDGEMENTS

The helpful discussion and assistance of Steven Blostein on two-view motion analysis are gratefully acknowledged.

The authors also wish to thank William Hoff who carefully read an early version of this paper and made valuable suggestions.

This research was supported by the National Science Foundation under grants ECS-83-52408 and ECS-83-19509.

REFERENCES

[1] G. Adiv, Determining three-dimensional motion and structure from optical flow generated by several moving objects, IEEE Trans. on Pattern Recognition and Machine Intelligence, vol. 7, no. 4, pp 348-401, July 1985

[2] O. Bottema and B. Roth, Theoretical kinematics, North-Holland, 1979

[3] W. L. Brogan, Modern control theory, Prentice-Hall, 1982

[4] T. J. Broida and R. Chellappa, Estimation of object motion parameters from noisy images, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 8, no.1, pp 90-99, Jan. 1986

[5] B. A. Conway, J. E. Tulinglowski, and P. D. Webber, Dynamics of remote orbital capture, Proc. AAS/AIAA Astrodynamics Specialist Conference, 1983

[6] J. Q. Fang and T. S. Huang, Some experiments on estimating The 3-D motion parameters of a rigid body from two consecutive image frames, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 6, no. 5, pp 547-554, Sept. 1984

[7] O. D. Faugeras and M. Hebert, A 3-D recognition and positioning algorithm using geometrical matching between primitive surfaces, Proc. 8th Int Joint Conference Artificial Intelligence, 1983, pp 996-1002

[8] G. R. Fowles, Analytical mechanics, Holt, Rinehart and Winston, 1977

[9] T. S. Huang, Motion analysis, to appear in AI Encyclopedia, Wiley, 1986

[10] T. S. Huang and S. D. Blostein, Robust algorithm for motion estimation based on two sequential stereo image pairs, Proc. Conf. Computer Vision and Pattern Recognition, 1985, pp 518-523

[11] M. H. Kaplan and A. A. Nadkarni, Control and stability problems of remote orbital capture, Mechanism and Machine Theory, vol. 12, 1977, pp 57-64

[12] C. L. Lawson and R. J. Hanson, Solving least squares problems, Prentice-Hall, 1974

[13] H. C. Longuet-Higgins, A computer program for reconstructing a scene from two projections, Nature, vol. 293, Sept. 1981, pp 133-135

[14] W. D. Macmillan, Dynamics of rigid bodies, Mcgraw-Hill, 1936

[15] R. Y. Tsai and T. S. Huang, Estimating 3-D motion parameters of a rigid planar patch, IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 29, No. 6, Dec. 1981, pp 1147-1152

[16] R. Y. Tsai and T. S. Huang, and W. L. Zhu, Estimating 3-D motion parameters of a rigid planar patch, II: singular value decomposition, IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 30, No. 4, Aug. 1982, pp 525-534

[17] R. Y. Tsai and T. S. Huang, Uniqueness and estimation of 3-D motion parameters of rigid bodies with curved surfaces, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 6, No.1, Jan. 1984, pp 13-27

[18] S. Ullman, The interpretation of visual motion, Massachusetts Institute Technology, 1979

[19] A. M. Waxman and J. H. Duncan, Binocular Image Flows, Proc. Image Understanding Workshop, 1985, pp 340-347

[20] A. M. Waxman and S. Ullman, Surface structure and 3-D motion from image flow: A kinematic analysis. Tech. Report 24. College Park, MD: University of Maryland, Center for Automation Research.

[21] Y. Yasumoto and G. Medioni, Experiments in estimation of 3-D motion parameters from a sequence of image frames, Proc. Computer Vision and Pattern Recognition, 1985, pp 89-94

[22] B. L. Yen and T. S. Huang, Determining 3-D motion and structure of a rigid body using straight line correspondences, Image Sequence Processing and Dynamic Scene Analysis, Springer-Verlag: Heidelberg, 1983

[23] X. Zhuang and R. M. Haralick, Rigid body motion and optic flow image, Proc. 1st Conf. on Artificial Intelligence Application, 1984, pp 366-375

[24] X. Zhuang and R. H. Haralick, Two view motion analysis, Proc. Int. Conf. on Acoustics, Speech, and Signal Processing, 1985