3-D Motion Estimation, Understanding, and Prediction from Noisy Image Sequences

JUYANG WENG, STUDENT MEMBER, IEEE, THOMAS S. HUANG, FELLOW, IEEE, AND NARENDRA AHUJA, SENIOR MEMBER, IEEE

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, i.e., the algorithm consists of solving simultaneous linear equations. 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.

Index Terms—**Computer vision, dynamic model, image sequence analysis, motion, motion estimation, motion prediction, motion understanding.**

I. 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 [25]. 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 translation and rotational 2-D motion and the structure of the object (the 3-D coordinates of the object points in the object-centered coordinate frame) is known. Yasumoto and Medioni [28] also assume the motion to be constant through the sequence and estimate, through a search in the solution space, the parameters of the assumed constant motion from image sequence. In the field of astrodynamics, the dynamic information about the object is required to be known. For instance, the principal moments of inertia and the structures of objects are required [8], [16]. Because the Lagrange equations of rigid body motion are nonlinear [12], [20], numerical methods are necessary to solve the dynamics problems [8], [16].

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 parameter derived, we can make extrapolations and interpolations

Manuscript received March 31, 1986; revised December 15, 1986. Recommended for acceptance by W. B. Thompson. This work was supported by the National Science Foundation under Grants ECS-83-52408 and ECS-83-19509.

The authors are with the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801.

IEEE Log Number 8613727.

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 object would be if it is missing from a image subsequence, 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 to 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, respectively. 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 [10], [19], [22], [23], [24], [31] 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 [19], and Tsai and Huang [24]. Line correspondences can also be used to solve the problem (Yen and Huang [29]). An alternative approach is to compute the optical flow field and then estimate motion parameters from optical flow [1], [26], [27], [30]. Zhuang and Haralick [30] 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 the ranges of object points. An algorithm to find the rotation and translation given 3-D correspondences has been discussed in Huang and Blostein [15]. Shuster [21], Faugeras and Herbert [11] propose the algorithm to find the least-squares solution of the motion parameters from 3-D coodinates in the presence of noise.

The problems of matching feature points in 2-D and 3-D have been discussed by several researchers. For monocular vision, Dreschler and Nagel [9] propose feature tracking from monocular image sequences. In binocular vision, the matching process may involve two steps. One is matching between two images taken at the same time instant (i.e., stereo matching) to derive 3-D coordinates of the features (e.g., Grimson [13], Hoff and Ahuja [14]). The other is matching these 3-D coordinates of features obtained at different time instants (e.g., Lin, Lee and Huang [18], Chen and Huang [7]). (Alternatively, the 3-D coordinates of the features may be obtained directly using range finders.)

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 II 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 III. Section III also deals with the local understanding, prediction, and recovering of the motion. Some particular properties of monocular vision are discussed in Section IV. Section V gives the results of simulations. Section VI presents a summary.

II. The LCAM Model

This section consists of four subsections. Subsection A deals with the general motion of a rigid body in 3-D. Sub-

section B 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 C discusses the solution of the coefficient equations. The relationship between the continuous precession of the LCAM model and the discrete two-view motion is investigated in Subsection D.

A. Motion of a Rigid Body in 3-D

All external forces acting on a body can be reduced to a total force $\boldsymbol{F}$ acting on a suitable point Q, and a total applied torquer $\boldsymbol{N}$ 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. That point may move with the supports. Letting m be the mass of the body, the motion of the center of mass is given by

$$\boldsymbol{F} = \frac{d}{dt}(m\boldsymbol{V}). \tag{1}$$

Let $\boldsymbol{L}$ be the angular momentum of the body. The torque $\boldsymbol{N}$ and the angular momentum $\boldsymbol{L}$ satisfy [12], [20]:

$$\boldsymbol{N} = \frac{d\boldsymbol{L}}{dt}. \tag{2}$$

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 rotation center Q will be discussed in the next subsection.

In matrix notation, the angular momentum $\boldsymbol{L}$ 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

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

$$I_{zz} = \int (x^2 + y^2)\, dm \qquad I_{zx} = I_{xz} = -\int xz\, dm$$

$$I_{yx} = I_{xy} = -\int xy\, dm \qquad I_{zy} = I_{yz} = -\int zy\, dm$$

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 [12], [20], the inertia tensor I takes the diagonal form:

$$I = \begin{bmatrix} I_{xx} & 0 & 0 \\ 0 & I_{yy} & 0 \\ 0 & 0 & I_{zz} \end{bmatrix}. \tag{3}$$

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

$$N_x = I_{xx}\dot{\omega}_x + \omega_y\omega_z(I_{zz} - I_{yy})$$

$$N_y = I_{yy}\dot{\omega}_y + \omega_z\omega_x(I_{xx} - I_{zz})$$

$$N_z = I_{zz}\dot{\omega}_z + \omega_x\omega_y(I_{yy} - I_{xx})$$

where $(N_x, N_y, N_z) = \boldsymbol{N}$. 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} in (3) are equal. Cylinders and disks are such examples. Most satellites are also symmetrical or almost symmetrical in this sense.

Under the above two simplifications, Euler's equations are integrable [12], [20]. The motion is such that the body rotates about its axis of symmetry $\boldsymbol{m}$, and at the same time the axis rotates about a spatially fixed $\boldsymbol{l}$. 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 $\boldsymbol{l}$, $\boldsymbol{\omega}_m$ be the angular velocity at which the rolling cone rotates about its own axis of symmetry $\boldsymbol{m}$. 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 $\boldsymbol{m}$ is parallel to $\boldsymbol{l}$. Then $\boldsymbol{\omega}$ is also parallel to $\boldsymbol{l}$. So the instantaneous rotation axis does not change its orientation in motion. This type of motion is called motion without precession.

In the rest of this paper column vectors and matrices

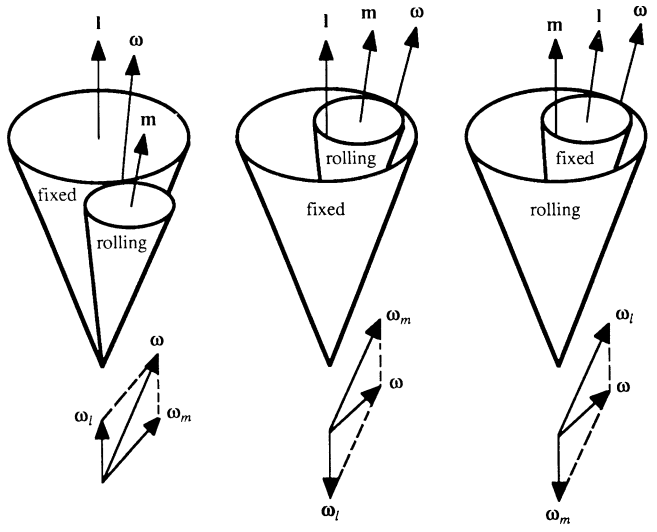


Fig. 1. The precessional motion of a torque-free symmetrical rigid body.

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, $i = 0, 1, \cdots$.

B. Motion of Rotation Center

The location of rotation center $Q(t)$ changes with time. Assume the trajectory of the rotation center is smooth, or specifically, it can be expanded into a Taylor series:

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

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 $(1/j!)\, Q^{(j)}(0) = b_{j+1}$, $j = 0, 1, 2, \cdots$, we have

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

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

$$Q_i = b_1 + cb_2 i + c^2 b_3 i^2 + \cdots + c^{k-1} b_i i^{k-1}. \quad (6)$$

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

$$Q_i = a_1 + a_2 i + a_3 i^2 + \cdots + a_k i^{k-1}. \quad (7)$$

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, (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 location 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 refer to 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 correspondence are needed to uniquely determine the rotation and the translation direction of the two-view motion using linear algorithms [24], [31]. 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 any 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

$$P_1 = R_1 P_0 + T_1 \quad (8)$$

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 (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 to (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 shown 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 Section II-A, 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.

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

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

or

$$-R_1Q_0 + Q_1 = T_1. \tag{11}$$

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

$$\begin{aligned} -R_1Q_0 + Q_1 &= T_1 \\ -R_2Q_1 + Q_2 &= T_2 \\ &\cdots \\ -R_fQ_{f-1} + Q_f &= T_f. \end{aligned} \tag{12}$$

Equations (12) give the relationship among the locations of the rotation center, the two-view rotation matrices, and the two-view translation vectors.

Substituting (7) into (12), we get

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

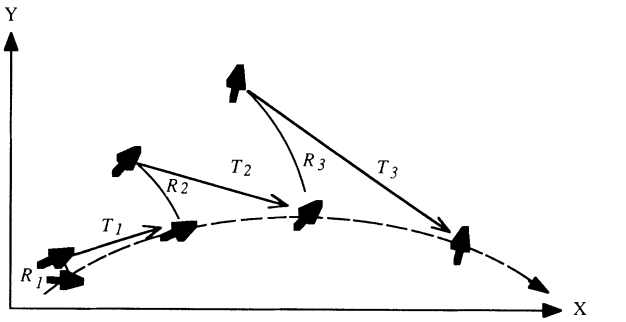


Fig. 2. Trajectory described by R_i and T_i if rotation axis is always located at origin.

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). Let the element of D at ith row and jth column be 3 by 3 matrix d_{ij}, i.e., $D = [d_{ij}]_{f \times k}$. We have

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

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).

C. Solutions of Coefficient Equation

Let $f = k$ in (13); 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):

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

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 has 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 the coefficient equations, let $f = k$. Define S_k^j to be a 3 by 3 matrix:

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

$$j = 0, 1, 2, \cdots, k.$$

Define number u_{ij}:

$$u_{ij} = \sum_{m=1}^{i} (-1)^{i-m} m^{j-1} \binom{i}{m}$$

$$j = i + 1, i + 2, \cdots, k.$$

Then

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

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

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

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

$$\cdots$$

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

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

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, $\cdots$, and last equations in Theorem 1. So if S_k^0 is not singular, the solution is unique.

Theorem 2: In the case of rotation without precession, letting w be any column vector parallel to the rotation axes, then,

$$S_k^0 w = 0 \quad (18)$$

and for any vector a

$$(S_k^0 a) \cdot w = 0. \quad (19)$$

Proof: See Appendix 2.

From Theorem 1, we have

$$S_k^0 a_1 = -\sum_{l=1}^{k} S_k^l T_l. \quad (20)$$

In the case of rotation without precession, (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 jth degree polynomial with $j \leq k - 1$ 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 (20). If the real trajectory is a jth degree polynomial with $j \leq k - 1$, then (20) has a solution by our derivation of (20). Since (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 a 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 (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 $S_{k-1}^0 a_1 = S_{k-1}^0 \hat{a}_1$, $S_{k-2}^0 a_1 = S_{k-2}^0 \hat{a}_1$, $\cdots$, $S_1^0 a_2 = S_1^0 \hat{a}_1$. By the equations given by Theorem 1, the unknowns a_k, a_{k-1}, $\cdots$, a_2 are determined without knowing the undetermined number 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 a zero matrix. The first three columns of D are zero, a_1 cannot 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 (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.

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 cannot be determined by coefficient equations. a_1 can be chosen to be the position of any object point at time t_0. Then (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. The algorithms in [21] and [11] can be used for the least-squares estimation of two-view motion parameters. To use overdetermination based on a large number of frames, we left $f > k$ 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 $f > k$, (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 (13) be a nonzero 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 this 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 [17]. 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.

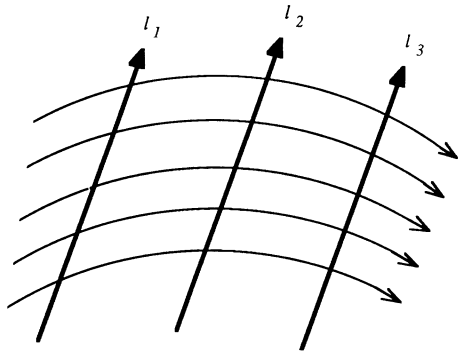


Fig. 3. The possible trajectories of rotation centers when rotation axes are parallel.

D. 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 Section II-A, a precession can be considered as the motion of 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 at time t_2. 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 the axis of the fixed cone, l, or vice versa.

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

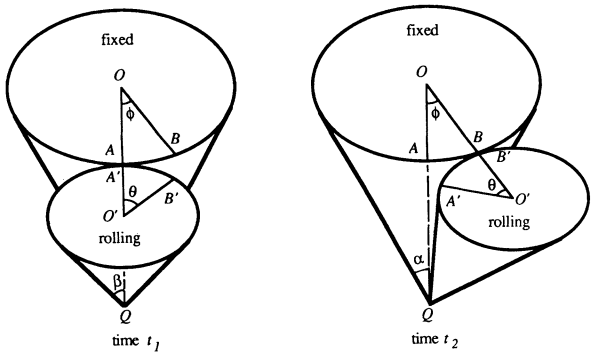


Fig. 4. The relation between rotation angles θ and ϕ.

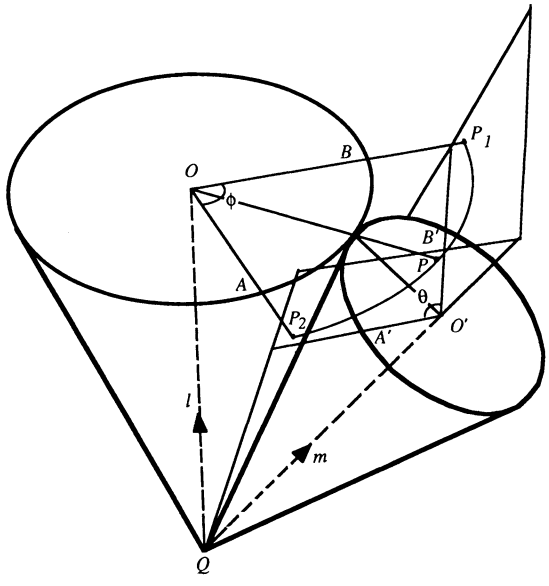


Fig. 5. Finding fixed points for two-view rotation.

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 (slipping along the fixed cone) 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 (slipping on the fixed cone) about its own axis m by angle θ. Point P now reaches position P_2. Finally, rotate the rolling cone (slipping along the fixed cone) 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 n 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.

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 frequency* ϕ to be the angular frequency at which the symmetrical axis of rolling cone rotates about the precession axis, define the *ith 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 (see Fig. 6).

From image sequences we find estimates of two-view motion parameters. They are the *ith 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 *ith two-view rotation angle* ψ_i and *ith two-view translation vector* T_i. Fig. 6 shows the precession parameters of continuous motion 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

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

$$R(m_i, \theta)\, R(l, \phi) = R(n_i, \psi_i). \quad (26)$$

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

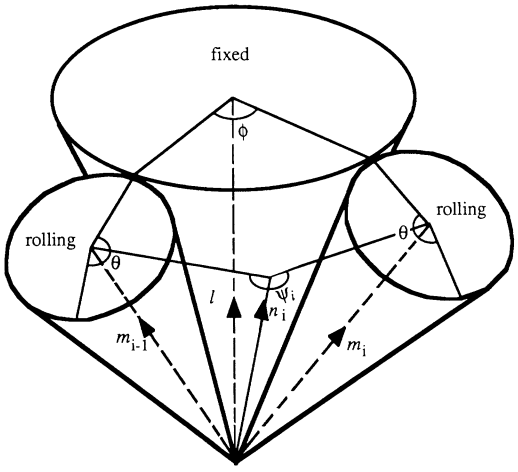


Fig. 6. Parameters of continuous precession and discrete two-view motion.

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 (25). Similarly, if we change the order of these two rotational components we get (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

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

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:

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

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 (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]:

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

$$l = \frac{[r_{32} - r_{23},\, r_{13} - r_{31},\, r_{21} - r_{12}]^t}{\left\| [r_{32} - r_{23},\, r_{13} - r_{31},\, {}_{21} - r_{12}] \right\|}. \tag{33}$$

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 (23).

III. Estimation, Understanding, and Prediction

The LCAM model is applied to subsequences of the images successively. The parameters of the model are estimated for every (overlapping) subsequence. The estimated model parameters can then be used to 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 of 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 pa-

rameter 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 the cameras and the viewing angle covering the object generally determined 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

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

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 $A^t A$. (See for example, [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 the 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 mode, the f + 1st two-view rotation axis vector n_{f+1} is $R(l, (f - i + 1)\phi)n_i$ for any $1 \le i \le f$. 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:

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

From (9) 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 $p \ge 2$ frames by using the following equations successively.

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

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

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 orientations of the object as well as the locations of the feature points on the object can be recovered by an 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.

IV. Monocular Vision

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

Let us assume a point P is located at $P_i = [x_i, y_i, z_i]^t$ at time t_i, $i = 0, 1, 2, \cdots$. Without loss of generality, we assume the focal length is unity. The image coordinates of P_i are:

$$(x_i/z_i, y_i/z_i). \tag{44}$$

Let the depth (z component) of P_0 be an unknown constant c, i.e., $z_0 = c$. Define a *c-scaled* version of P_i:

$$\tilde{P}_i = \frac{1}{c} P_i. \tag{45}$$

Once $\tilde{P}_i$ is determined, we have determined the 3-D coordinates of P_i up to a global scale factor c. Applying the

motion prediction algorithm using $\tilde{P}_i$ as 3-D coordinates instead of P_i (8 or more *c-scaled* versions of points are needed to estimate the motion parameters [24], [31]), we can predict the 3-D position of the point P up to a global scale factor c. The image coordinates are still unambiguously predicted since the global scale factor c is canceled out in image coordinates [see (44)]. In fact it is true that we can determine $\tilde{P}_i$, $i = 1, 2, \cdots$ as we show below.

Let the two-view motion from frame F_{i-1} to F_i be represented by rotation matrix R_i and translation vector T_i; we have

$$P_i = R_i P_{i-1} + T_i. \tag{46}$$

Assuming $\| T_i \| \neq 0$, we get

$$\frac{P_i}{\| T_i \|} = R_i \frac{P_{i-1}}{\| T_i \|} + \frac{T_i}{\| T_i \|}. \tag{47}$$

Using eight or more point correspondences between F_{i-1} and F_i, we determine R_i, the translation direction $\hat{T}_i = T_i / \| T_i \|$ and the relative 3-D coordinates of the points P_{i2} and P_{i1} satisfying

$$P_{i2} = \frac{P_i}{\| T_i \|}, \qquad P_{i1} = \frac{P_{i-1}}{\| T_i \|} \tag{48}$$

(47) becomes

$$P_{i2} = R_i P_{i1} + \hat{T}_i. \tag{49}$$

Using (45) and (48) and letting $i = 1$, we have

$$c\tilde{P}_0 = P_0 = \| T_1 \| P_{11} \tag{50}$$

or,

$$\| T_1 \| = c \frac{\| \tilde{P}_0 \|}{\| P_{11} \|}. \tag{51}$$

From (44) and (45) we know $\tilde{P}_0 = [x_0/z_0, y_0/z_0, 1]^t$. So $\tilde{P}_0$ can be determined by the image coordinates of P_0. T_1 then is determined up to a scale factor c in (51). From (48) and (51) we have

$$P_1 = \| T_1 \| P_{12} = c \frac{\| \tilde{P}_0 \|}{\| P_{11} \|} P_{12}. \tag{52}$$

Thus using (45), we get

$$\tilde{P}_1 = \frac{\| \tilde{P}_0 \|}{\| P_{11} \|} P_{12}. \tag{53}$$

Similarly, we can determine $\tilde{P}_2, \tilde{P}_3, \cdots$, successively.

One point should be mentioned here. In the binocular case the set of corresponding points may be different from different consecutive image frame pairs. This means that some points 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 the above discussion we implicitly assume the point P is visible through all the image frames. However, it is not necessary.

For example, assume P is not visible in frame F_2 and another point Q is visible in F_0, F_1, F_2. Q can be used to convey the constant c further when P is invisible. Letting Q replace P in (46) and using (51) we have

$$Q_1 - R_1 Q_0 = \| T_1 \| \hat{T}_1 = c \frac{\| \tilde{P}_0 \|}{\| P_{11} \|} \hat{T}_1. \tag{54}$$

Letting $Q_i = [u_i, v_i, w_i]^t$, $i = 0, 1, 2$, and using (54), we have

$$w_1 \begin{bmatrix} u_1/w_1 \\ v_1/w_1 \\ 1 \end{bmatrix} - R_1 w_0 \begin{bmatrix} u_0/w_0 \\ v_0/w_0 \\ 1 \end{bmatrix} = c \frac{\| \tilde{P}_0 \|}{\| P_{11} \|} \hat{T}_1. \tag{55}$$

In (54), the two column vectors to the left side are determined by the image coordinates of Q_1 and Q_0, respectively. The unknowns in (55) are numbers w_1, w_0, and c. Equation (55) is a set of three linear equations for these unknowns. The solutions of (55), with w_1 and w_0 as unknowns and c as a parameter, are proportional to the parameter c, i.e., we can find $\tilde{w}_1$ such that $w_1 = c\tilde{w}_1$. Then, letting $\tilde{Q}_1 = \tilde{w}_1 [u_1/w_1, y_1/w_1, 1]^t$, we have $Q_1 = c\tilde{Q}_1$. As for determining $\| T_2 \|$, $\tilde{Q}_1$ here can take the role of $\tilde{P}_1$ using the following equation similar to (51):

$$\| T_2 \| = c \frac{\| \tilde{Q}_1 \|}{\| Q_{21} \|} \tag{56}$$

where $\| Q_{21} \|$ is determined in two-view motion analysis [see (49)].

In summary, the *c-scaled* version of the translation vector can be determined from the *c-scaled* version of the 3-D position vector of a point at the time before two-view motion; the *c-scaled* version of the 3-D position vector of every 3-D point (used in the point correspondences) before and after the two-view motion can be determined from the *c-scaled* version of the translation vector. So, we choose the depth of any point at time t_0 as the unknown scalar c. All the 3-D points and the translation vectors are then determined up to that single constant c. It is easy to see from the above discussion that as long as every three consecutive frames have at least one common visible point, all the 3-D coordinates of the points used to estimate the motion parameters as well as the translation vectors can be determined up to a global scale factor c.

V. 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 the coefficient vectors a_i in (7). By Theorem 3, the rotation

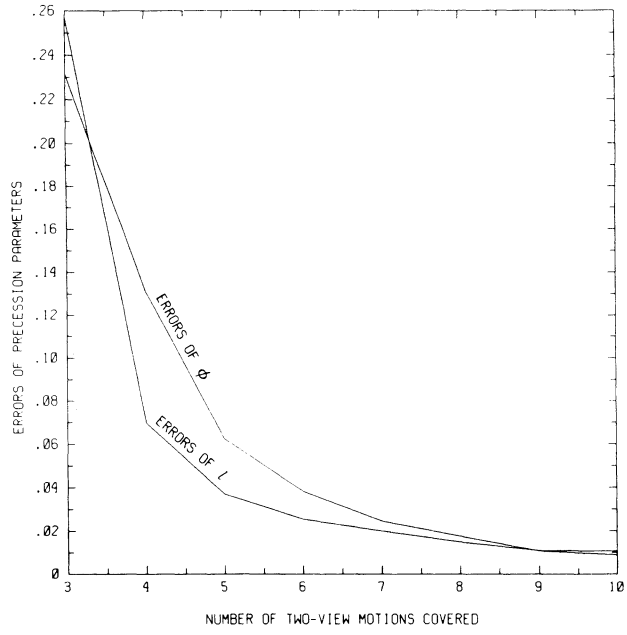


Fig. 7. Errors of l (precession vector) and ϕ (precession angular frequency) versus the number of two-view motions covered. Number of point correspondences: 3. Frame predicted: next. Image resolution: 256 by 256.

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 10 cm. Simulated cameras are 100 cm away from the object. The viewing lines from the two cameras to the object form an angle of 45°. 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. The cube covered about one fourth of the area of the whole image. So, the actual image resolution is reduced by a factor of 2, i.e., if the image resolution is denoted as 64 by 64 in the following figures, actually about 32 by 32 resolution is used.

The errors in the image coordinates of the feature points arise from several sources such as digitization noise, corner detector errors and lens distortion errors, etc. In a well calibrated data acquisition system, lens distortion is well compensated and so could be ignored. The corner detector errors can be approximated as digitization noise. For example, an error range of plus or minus one pixel in the corner detector could be approximated by the reducing the image resolution by a factor of two. The resolution of 64 by 64 approximates the case of a 512 by 512 image with the additional corner detector error of range ± 4 pixels. For the simulations, the object covers only about one fourth of the images. So, if the object covered the whole image, the errors shown in the simulation data should correspond to images at half the resolutions.

So in the simulation the noise is represented by the choice of image resolution, i.e., noise is due to the spatial quantization of the image plane. An increase in noise can be represented by a decrease in image resolution. Because of digitization errors, the set of points ceases to satisfy the rigid body constraint. Therefore, to find the best two-view motion parameters, a least-squares solution is obtained [21], [11].

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: $n_1 = (1/\sqrt{17})\,[1, 0, 4]^t$. Coefficient vectors: $a_1 = [-2, -3, -1]^t$, $a_2 = [0.5, 0.5, 0.25]^t$, $a_3 = [0.005, 0.005, 0.0025]^t$. 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 vector of the estimated and the real unit vectors. The errors approach zero very fast as the number of two-view motions 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 the prediction errors. The relative errors are defined as the prediction errors divided by the length of the diagonal of the cube. The results of the predicting more frames are presented in Fig. 8. Fig. 9 shows the mean relative prediction errors at time t_5 versus image resolution for diffferent numbers of point correspondences. Fig. 10 shows the mean relative prediction errors versus

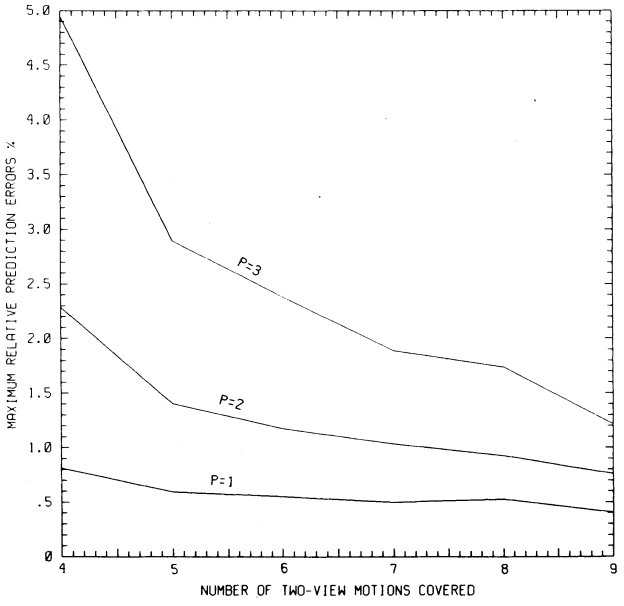


Fig. 8. Maximum relative prediction errors (percent) versus the number of two-view motions covered for different numbers of frames predicted (p). Frame predicted: next pth frame. Number of point correspondences: 3. Image resolution: 512 by 512. Degree of polynomial $k - 1$ in (7): 2.

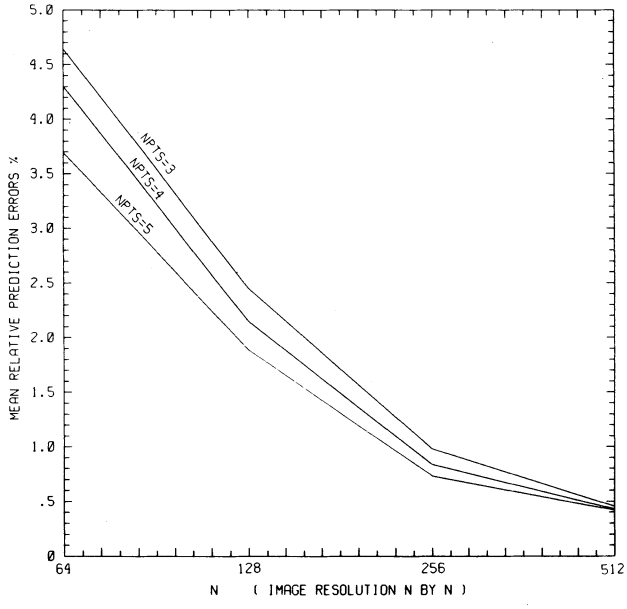


Fig. 9. Mean relative prediction errors (percent) versus the image resolution (n by n) for different numbers of point correspondences (npts). Frame predicted: next frame. Number of two-view motions covered: 5. Degree of polynomial $k - 1$ in (7): 2.

image resolution for different number of two-view motions covered. Fig. 11 gives the mean relative prediction errors versus the number of two-view motions covered for different numbers of point correspondences.

The 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 tips of the wings, stabilizers and rudder of the airplane. Since not all of these five tips are visible in all images, four currently visible

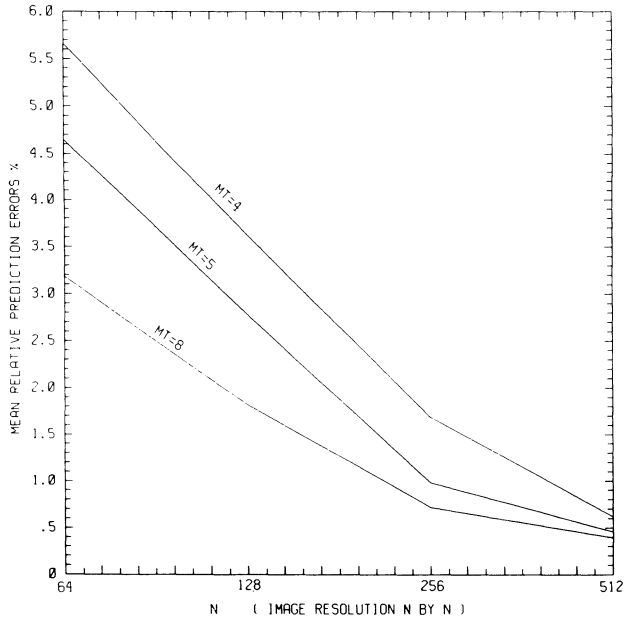


Fig. 10. Mean relative prediction errors (percent) versus image resolution (n by n) for different numbers of two-view motions covered (mt). Frame predicted: next frame. Number of point correspondences: 3. Degree of polynomial $k - 1$ in (7): 2.

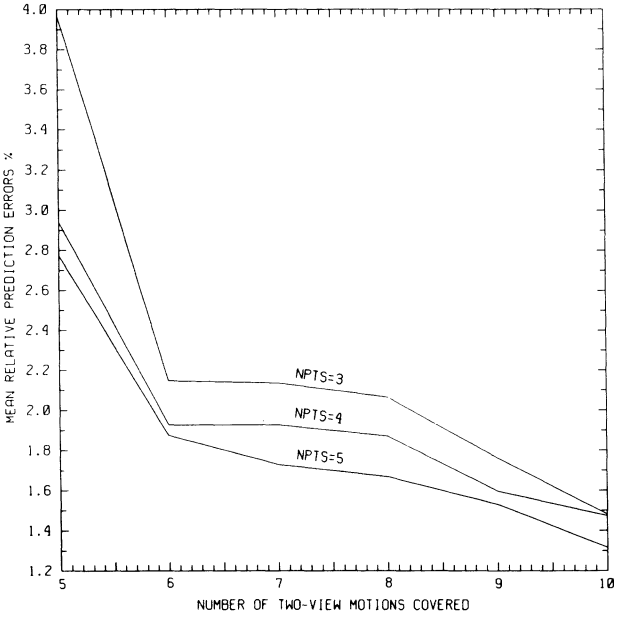


Fig. 11. Mean relative prediction errors (percent) versus the number of two-view motions covered for different numbers of point correspondences (npts). Frame predicted: next frame. Image resolution: 128 by 128. Degree of polynomial $k - 1$ in (7): 3.

tips are used as feature point sets through the image sequence. The number of point correspondences is then 4. The feature points and their correspondences are determined manually based on the gray tone of the digitized images. The corner detector error here is then the error of manual detection based on the gray tone of the pixels. The camera resolution is 512 by 512. 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. Compared with the

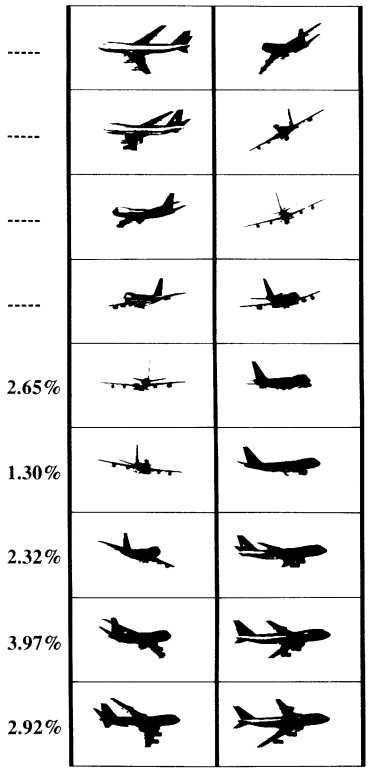


Fig. 12. Image frame sequence of a model airplane. Left column: left view; right column: right view. (From the fifth frame, the maximum relative prediction errors are shown to the left of the corresponding frame pairs.)

synthetic data case, the additional error source is lens distortion. The relative errors in this experiment are close to the synthetic data case.

VI. 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 discuss the former. 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, the 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.

Recently, Broida and Chellappa [5], [6] have proposed a framework to deal with more general 3-D motion than their early work [4]. One of the major differences between our work and their work is that our model is based on dynamics but those in [4]–[6] are not. It is assumed in [4], [6] that the complete geometrical structure of the object is known. For the cases where angular velocity varies with time (this is the case in our formulations and experiments) Broida and Chellappa [6] proposed numerical in-

tegration but neither specific algorithms nor experiments were given. The approach presented in our paper assumes a general precessional motion. The relatively weak assumption of a symmetrical object is a sufficient (but not necessary) condition for such a precessional motion. The symmetry assumption can be dropped as long as the motion can be modeled in the short term by the precessional motion discussed.

APPENDIX 1

Lemma 1: Define

$$\hat{S}_k^j = \sum_{l=0}^{k-j} (-1)^l \binom{k}{l} \prod_{m=j+2}^{k-l+1} R_m.$$

Let $\Pi_{m=i}^{j} R_m = I$, if $i > j$. Then

$$\hat{S}_k^0 R_1 - S_k^0 = S_{k+1}^0.$$

Proof:

$$\begin{aligned}
\hat{S}_k^0 R_1 - S_k^0 &= \sum_{l=0}^{k} (-1)^l \binom{k}{l} \prod_{m=1}^{k-l+1} R_m - \sum_{l=0}^{k} (-1)^l \binom{k}{l} \prod_{m=1}^{k-l} R_m \\
&= \binom{k}{0} \prod_{m=1}^{k+1} R_m + \sum_{l=1}^{k} (-1)^l \binom{k}{l} \prod_{m=1}^{k-l+1} R_m - \sum_{l=0}^{k-1} (-1)^l \binom{k}{l} \prod_{m=1}^{k-l} R_m - (-1)^k \binom{k}{k} I \\
&= \binom{k+1}{0} \prod_{m=1}^{k+l} R_m + \sum_{l=1}^{k} (-1)^l \binom{k}{l} \prod_{m-1}^{k-l+1} R_m \qquad \text{(A1)} \\
&\quad + \sum_{j=1}^{k} (-1)^j \binom{k}{j-1} \prod_{m=1}^{k-j+1} R_m + (-1)^{k+1} \binom{k+1}{k+1} I.
\end{aligned}$$

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

$$\begin{aligned}
\hat{S}_k^0 R_1 - S_k^0 &= \binom{k+1}{0} \prod_{m=1}^{k+1} R_m + \sum_{l=1}^{k} (-1)^l \binom{k+1}{l} \prod_{m=1}^{k-l+1} R_m + (-1)^{k+1} \binom{k+1}{k+1} I \\
&= \sum_{l=0}^{k+1} (-1)^l \binom{k+1}{l} \prod_{m=1}^{k-l+1} R_m \\
&= S_{k+1}^0.
\end{aligned}$$

□

Lemma 2: Let $k \geq 2$, $1 \leq m \leq k-1$. Then

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

$$\sum_{j=1}^{k} (-1)^{k-j} j^k \binom{k}{j} = k! \qquad \text{(A.3)}$$

Proof: From the binomial equation

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

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

$$\begin{aligned}
&(1+x)^{k-2} (k(k-1)x^2 + k(1+x)x) \\
&\quad = \sum_{j=1}^{k} j^2 \binom{k}{j} x^j.
\end{aligned}$$

Generally, doing this m times, $1 \leq m \leq k$, gives

$$\begin{aligned}
&(1+x)^{k-m} [k(k-1) \cdots (k-m+1) x^m \\
&\quad + (1+x) P(x)] = \sum_{j=1}^{k} j^m \binom{k}{j} x^j \qquad \text{(A.4)}
\end{aligned}$$

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

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

Multiplying both sides by $(-1)^m$ yields (A.2). For $m = k$, (A.4) gives

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

Multiplying both sides by $(-1)^k$ yields (A.3). □

Theorem 1: Let $f = k$ in coefficient equations. Define S_k^j to be a 3 by 3 matrix:

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

where $\prod_{m=i}^{j} R_m = I$, if $i > j$. Define number u_{ij}:

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

Then

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

Proof: We first prove

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

Let $c(t)$ be the position vector of the rotation center at time t, c_i be the position vector of the rotation center at time t_i. For conciseness, we define $T_0 = c_0$. (Note $c_0 = a_1$.) We have

$$c_0 = T_0$$
$$c_1 = R_1 c_0 + T_1 = R_1 T_0 + T_1$$
$$c_2 = R_2 c_1 + T_2 = R_2 R_1 T_0 + R_2 T_1 + T_2$$
$$\cdots$$

We get

$$c_i = \sum_{l=0}^{i} \prod_{m=l+1}^{i} R_m T_l \qquad 0 \le i \le k. \tag{A.6}$$

The $(k-1)$th degree vector polynomial that goes through c_i, $i = 1, 2, \cdots, c_k$, is

$$c(t) = \sum_{i=1}^{k} \frac{\prod_{\substack{1 \le j \le k \\ j \ne i}} (t - t_j)}{\prod_{\substack{1 \le j \le k \\ j \ne i}} (t_i - t_j)}.$$

Or, using (A.6)

$$c(t) = \sum_{i=1}^{k} \sum_{l=0}^{i} \frac{\prod_{\substack{1 \le j \le k \\ j \ne i}} (t - t_j)}{\prod_{\substack{1 \le j \le k \\ j \ne i}} (t_i - t_j)} \prod_{m=l+1}^{i} R_m T_l. \tag{A.7}$$

Since the time intervals are the same constant c, we have $t_i - t_j = c(i-j)$. Letting $t = t_0$ in (A.7), we get

$$c_0 = \sum_{i=1}^{k} \sum_{l=0}^{i} \frac{(-1)^{k-1} k!/i}{(-1)^{k-i}(i-1)!(k-i)!} \prod_{m=l+1}^{i}$$
$$R_m T_l = \sum_{i=1}^{k} \sum_{l=0}^{i} (-1)^{i-1} \binom{k}{i} \prod_{m=l+1}^{i} R_m T_l. \tag{A.8}$$

Since $T_0 = c_0 = a_1$, (A.8) gives

$$a_1 = \sum_{i=1}^{k} \sum_{l=1}^{i} (-1)^{i-1} \binom{k}{i} \prod_{m=l+1}^{i} R_m T_l + \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \prod_{m=1}^{i} R_m a_1.$$

Moving the term of a_1 to the left yields

$$\left(\sum_{i=0}^{k} (-1)^i \binom{k}{i} \prod_{m=1}^{i} R_m\right) a_1 = -\sum_{i=1}^{k} \sum_{l=1}^{i} (-1)^i \binom{k}{i} \prod_{m=l+1}^{i} R_m T_l. \tag{A.9}$$

Since $\sum_{i=1}^{k} \sum_{l=1}^{i}$ and $\sum_{l=1}^{k} \sum_{i=l}^{k}$ sum the same terms, we get

$$\left(\sum_{i=0}^{k} (-1)^i \binom{k}{i} \prod_{m=1}^{i} R_m\right) a_1 = -\sum_{l=1}^{k} \sum_{i=l}^{k} (-1)^i \binom{k}{i} \prod_{m=l+1}^{i} R_m T_l. \tag{A.10}$$

Multiplying both sides by $(-1)^k$ and letting $j = k - i$, we have

$$\left(\sum_{j=0}^{k} (-1)^j \binom{k}{j} \prod_{m=1}^{k-j} R_m\right) a_1 = -\sum_{l=1}^{k} \sum_{j=0}^{k-l} (-1)^j \binom{k}{j} \prod_{m=l+1}^{k-j} R_m T_l. \tag{A.11}$$

Using the definition of S_k^l (A.11) gives (A.5).

We now prove the rest of the equations. Let $2 \le i \le k-1$. Consider the first i equations in (13). Moving the terms of $a_{i+1}, \cdots, a_k$ to the right side, we get a set of i equations similar to (13) with the terms on the right side changed. Using the results in (A.5) for this set of equations to solving a_1, we have

$$S_i^0 a_1 = -\sum_{m=1}^{i} S_i^m \left(T_m - \sum_{j=i+1}^{k} d_{mj} a_j\right) = -\sum_{m=1}^{i} S_i^m T_m + \sum_{j=i+1}^{k} \left(\sum_{m=1}^{i} S_i^m d_{mj}\right) a_j \tag{A.12}$$

where

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

$$
\begin{aligned}
&= \sum_{m=1}^{i-1} m^{j-1} \sum_{l=0}^{i-m} (-1)^l \binom{i}{l} \prod_{n=m+1}^{i-l} R_n \\
&\quad - \sum_{m=2}^{i} (m-1;)^{j-1} \sum_{l=0}^{i-m} (-1)^l \binom{i}{l} \prod_{n=m}^{i-l} R_n \\
&\quad + i^{j-1} I \\
&= \sum_{m=1}^{i-1} m^{j-1} \sum_{l=0}^{i-m-1} (-1)^l \binom{i}{l} \prod_{n=m+1}^{i-l} R_n \\
&\quad + \sum_{m=1}^{i-1} (-1)^{i-m} m^{j-1} \binom{i}{i-m} I \\
&\quad - \sum_{v=1}^{i-1} v^{j-1} \sum_{l=0}^{i-v-1} (-1)^l \binom{i}{l} \prod_{n=v+1}^{i-l} R_n \\
&\quad + i^{j-1} I \\
&= \left(\sum_{m=1}^{i-1} (-1)^{i-m} m^{j-1} \binom{i}{m} + i^{j-1} \right) I \\
&= \left(\sum_{m=1}^{i} (-1)^{i-m} m^{j-1} \binom{i}{m} \right) I \\
&= u_{ij} I.
\end{aligned}
$$

So, (A.12) yields

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

or,

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

By Lemma 2, $u_{i(i+1)} = i!$ So we have

$$a_{i+1} = \frac{1}{i!} \left(\sum_{m=1}^{i} S_i^m T_m + S_i^0 a_1 - \sum_{j=i+2}^{k} u_{ij} a_j \right). \tag{A.13}$$

Letting $i = k-1, k-2, \cdots, 2, 1$, we get the equations of Theorem 1 presented in Section II-C. □

Appendix 2

Theorem 2: In the case of rotation without precession, letting w be any vector parallel to the rotation axes, then,

$$S_k^0 w = 0 \tag{A.15}$$

and for any vector a

$$(S_k^0 a) \cdot w = 0. \tag{A.16}$$

Proof: We prove (A.16). The proof of (A.15) 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 (see Lemma 1) in Appendix 1)

$$(\hat{S}_k^0 b) \cdot w = 0. \tag{A.17}$$

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 $(\hat{S}_1^0 b) \cdot w = 0$.

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

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

Similarly (A.17) holds for $k = n + 1$. □

Appendix 3

Let R_1 and R_2 be the rotation matrices with rotation axes n_1 and n_2, rotation angle θ_1 and θ_2, respectively. $A = R_2 R_1 - 2R_1 + I$ is singular if and only if n_1 and n_2 are parallel or $\theta_1 = 0$, or $\theta_2 = 0$.

Proof: Without loss of generality, let n_1 and n_2 be unit vectors. Since the points on the rotation axis do not change the positions after motion, we have $R_i n_i = n_i$, $i = 1, 2$.

If $\theta_1 = 0$, then $R_1 = I$ and we have $An_2 = R_2 n_2 - 2n_2 + n_2 = 0$. So A is singular if $\theta_1 = 0$. Similarly if $\theta_2 = 0$, A is singular. If n_1 and n_2 are parallel, we have $R_2 n_1 = n_1 = R_1 n_1$. Thus $An_1 = R_2 n_1 - 2R_1 n_1 + n_1 = 0$, and so A is singular.

Conversely, assume A is singular. There is a unit vector x such that $Ax = (R_2 R_1 - 2R_1 + I)\, x = 0$. Or,

$$R_2 R_1 x = 2R_1 x - x \tag{A.18}$$

There exists a unit vector $\hat{x}$ which is perpendicular to x such that $R_1 x$ can be represented by $R_1 x = ax + b\hat{x}$, where a and b are real numbers. Since $R_1 x$ is a unit vector, $a^2 + b^2 = 1$. From (A.18) we get

$$
\begin{aligned}
1 &= \|x\|^2 = \|R_2 R_1 x\|^2 = \|2R_1 x - x\|^2 \\
&= \|(2a - 1)\, x + 2b\hat{x}\|^2 \\
&= (2a - 1)^2 + (2b)^2 \\
&= 4a^2 - 4a + 1 + 4b^2 = 5 - 4a
\end{aligned}
$$

Thus $a = 1$. So $b = 0$. We have

$$R_1 x = x \tag{A.19}$$

Using (A.18) and (A.19) yields

$$R_2 x = R_2 R_1 x = 2R_1 x - x = 2x - x = x. \tag{A.20}$$

Unless $\theta_1 = 0$ or $\theta_2 = 0$, (A.19) and (A.20) imply that x is parallel to the rotation axes of R_1 and R_2. So, n_1 and n_2 are parallel unless $\theta_1 = 0$ or $\theta_2 = 0$. □

Acknowledgment

The authors would like to thank Y. S. Shim for the discussions on shortening the proofs presented in Appendix 1 and Appendix 3.

REFERENCES

[1] G. Adiv, "Determining three-dimensional motion and structure from optical flow generated by several moving objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, pp. 348–401, 1985.

[2] O. Bottema and B. Roth, *Theoretical Kinematics*. New York: North-Holland, 1979.

[3] W. L. Brogan, *Modern Control Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

[4] T. J. Broida and R. Chellappa, "Estimation of object motion parameters from noisy images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 90–99, 1986.

[5] ——, "Kinematics and structure of a rigid object from a sequence of noisy images," in *Proc. IEEE Workshop on Motion: Representation and Analysis*, Charleston, SC, May 1986, pp. 95–100.

[6] ——, "Kinematics and structure of a rigid object from a sequence of noisy images: A batch approach," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Miami Beach, FL, June 1986, pp. 176–182.

[7] H. H. Chen and T. S. Huang, "Maximal matching of two three-dimensional point sets," in *Proc. 8th Int. Conf. Pattern Recognition*, Paris, France, Oct. 1986.

[8] B. A. Conway, J. E. Tulinglowski, and P. D. Webber, "Dynamics of remote orbital capture," in *Proc. AAS/AIAA Astrodynamics Specialist Conf.*, 1983.

[9] L. Dreschler and H.-H. Nagel, "Volumetric model and 3-D trajectory of a moving car derived from monocular TV frame sequences of a street scene," *Comput. Graphics Image Processing*, vol. 20, pp. 199–228, 1982.

[10] 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. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 547–554, 1984.

[11] O. D. Faugeras and M. Hebert," A 3-D recognition and positioning algorithm using geometrical matching between primitive surfaces," in *Proc. 8th Int. Joint Conf. Artificial Intell.*, Karlsruhe, West Germany, Aug. 1983, pp. 996–1002.

[12] G. R. Fowles, *Advanced Mechanics*, 3rd ed. New York: Holt, Rinehart and Winston, 1977.

[13] W. E. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, pp. 17–34, 1985.

[14] W. Hoff and N. Ahuja, "Depth from stereo," in *Proc. 4th Scandinavian Conf. Image Analysis*, Trondheim, Norway, June 1985.

[15] T. S. Huang and S. D. Blostein, "Robust algorithm for motion estimation based on two sequential stereo image pairs," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, 1985, pp. 518–523.

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

[17] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

[18] Z. Lin, H. Lee, and T. S. Huang, "Finding 3-D point correspondences in motion estimation," in *Proc. 8th Int. Conf. Pattern Recognition*, Paris, France, Oct. 1986.

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

[20] W. D. Macmillan, *Dynamics of Rigid Bodies*. New York: McGraw-Hill, 1936.

[21] M. D. Shuster, "Approximate algorithms for fast optimal attitude computation," in *Proc. AIAA Guidance and Control Specialist Conf.*, Palo Alto, CA, Aug. 1978, pp. 88–95.

[22] R. Y. Tsai and T. S. Huang, "Estimating 3-D motion parameters of a rigid planar patch," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 1147–1152, 1981.

[23] R. Y. Tsai, T. S. Huang, and W. L. Zhu, "Estimating 3-D motion parameters of a rigid planar patch, II: Singular value decomposition." *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 525–534, 1982.

[24] R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of 3-D motion parameters of rigid bodies with curved surfaces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 13–27, 1984.

[25] S. Ullman, *The Interpretation of Visual Motion*. Cambridge, MA: M.I.T. Press, 1979.

[26] A. M. Waxman and J. H. Duncan, "Binocular image flows," in *Proc. Image Understanding Workshop*, Miami, FL, 1985, pp. 340–347.

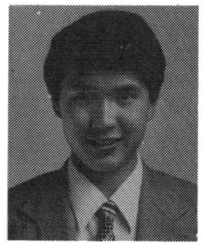
[27] A. M. Waxman and S. Ullman, "Surface structure and 3-D motion from image flow: A kinematic analysis," Center for Automation Research, Univ. Maryland, College Park, Tech. Rep. 24.

[28] Y. Yasumoto and G. Medioni, "Experiments in estimation of 3-D motion parameters from a sequence of image frames," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, 1985, pp. 89–94.

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

[30] X. Zhuang and R. M. Haralick, "Rigid body motion and optic flow image," in *Proc. IEEE 1st Conf. Artificial Intell. Application*, 1984, pp. 366–375.

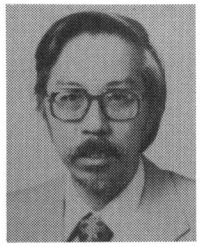
[31] X. Zhuang, T. S. Huang, and R. M. Haralick, "Two-view motion analysis: A unified algorithm," *J. Opt. Soc. Amer., A*, vol. 3, no. 9, pp. 1492–1500, Sept. 1986.



Juyang Weng (S'85) was born in Shanghai, China, on April 15, 1957. He received the B.S. degree in computer science from Fudan University, Shanghai, China, in 1982, and M.S. degree in computer science from the University of Illinois, Urbana-Champaign, in 1985.

Currently he is a research assistant at the Coordinated Science Laboratory, University of Illinois, Urbana-Champaign, and is working toward the Ph.D. degree in computer science. His current research interests are in computer vision, image sequence analysis, image processing, solid modeling and representation, and artificial intelligence.

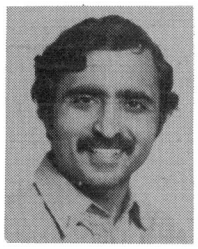
Mr. Weng is a member of Phi Kappa Phi.



Thomas S. Huang (S'61–M'63–SM'76–F'79) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, China, and the M.S. and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge.

He was on the Faculty of the Department of Electrical Engineering at M.I.T. from 1963 to 1973, and on the Faculty of the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purdue University from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, where he is now Professor of Electrical and Computer Engineering and Research Professor at the Coordinated Science Laboratory. During his sabattical leaves, he has worked at M.I.T. Lincoln Laboratory, IBM Thomas J. Watson Research Center, and the Rheinishes Landes Museum in Bonn, West Germany, and held visiting Professor positions at the Swiss Institutes of Technology in Zürich and Lausanne, the University of Hannover in West Germany, and INRS-Telecommunications of the University of Quebec in Montreal, Canada. He has served as a consultant to numerous industrial firms and government agencies both in the U.S. and abroad. His professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 10 books, and over 200 papers on network theory, digital filtering, image processing, and computer vision.

Dr. Huang is a Fellow of the Optical Society of America. He received a Guggenheim Fellowship (1971–1972), an A. V. Humboldt Foundation Senior U.S. Scientist Award (1976–1977), and a Fellowship from the Japan Society for the Promotion of Science (1986). He is an Editor of the international journal *Computer Vision, Graphics, and Image Processing*, Editor of the *Springer Series in Information Sciences*, published by Springer-Verlag, and Editor of the *Research Annual Series on Advances in Computer Vision and Image Processing* published by the JAI Press.



Narendra Ahuja (S'79–M'79–SM'85) received the B.E. degree with honors in electronics engineering from the Birla Institute of Technology and Science, Pilani, India, in 1972, the M.E. degree with distinction in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1974, and the Ph.D. degree in computer science from the University of Maryland, College Park, in 1979.

From 1974 to 1975 he was Scientific Officer in the Department of Electronics, Government of India, New Delhi. From 1975 to 1979 he was at the Computer Vision Laboratory, University of Maryland, College Park, as a Graduate Research Assistant (1975–1978), as a Faculty Research Assistant (1978–1979), and as a Research Associate (1979). In 1979 he joined the University of Illinois, Urbana-Champaign, where he is currently an Associate Professor in the Coordinated Science Laboratory and an Associate Professor in the Department of Electrical and Computer Engineering. His research interests are in computer vision, artificial intelligence, image processing, architectures for vision, and parallel algorithms.

Dr. Ahuja has received the University Scholar Award (1985) and the Presidential Young Investigator Award (1984). He is listed in *Who's Who in Technology Today* and *Who's Who in the Midwest*, and is a member of the Association for Computing Machinery and the American Association for Artificial Intelligence.