

Motion and Structure from Feature Correspondences: A Review

THOMAS S. HUANG, FELLOW, IEEE, AND ARUN N. NETRAVALI, FELLOW, IEEE

We present a review of algorithms and their performance for determining three-dimensional (3D) motion and structure of rigid objects when their corresponding features are known at different times or are viewed by different cameras. Three categories of problems are considered, depending upon whether the features are two- (2D) or three-dimensional (3D) and the type of correspondence: a) 3D to 3D (i.e., locations of corresponding features in 3D space are known at two different times), b) 2D to 3D (i.e., locations of features in 3D space and their projection on the camera plane are known), and c) 2D to 2D (i.e., projections of features on the camera plane are known at two different times). Features considered include points, straight lines, curved lines, and corners. Emphasis is on problem formulation, efficient algorithms for solution, existence and uniqueness of solutions, and sensitivity of solutions to noise in the observed data. Algorithms described have been used in a variety of applications. Some of these are: a) positioning and navigating 3D objects in a 3D world, b) camera calibration, i.e., determining location and orientation of a camera by observing 3D features whose location is known, c) estimating motion and structure of moving objects relative to a camera. We mention some of the mathematical techniques borrowed from algebraic geometry, projective geometry, and homotopy theory that are required to solve these problems, list unsolved problems, and give some directions for future research.

I. INTRODUCTION

Human beings are endowed with the ability to see, discern objects, estimate and understand motion, and navigate in the three-dimensional (3D) space. Incorporating such ability in machines has been a challenging task and has occupied scientists and engineers working in computer vision for a long time [1], [2]. In this paper we review a subset of this activity which deals with observing some features on the surface of an object in the environment at different points in time and using this information to derive 3D motion and structure of these objects. Depending on the nature of the features observed (two- or three-dimensional, points or lines on the surface of the object, etc.), different formulations and algorithms come into play. However, the underlying mathematics has much in common: all the different cases can be formulated in

such a way that they require solution of simultaneous transcendental, polynomial, or linear equations in multiple variables which represent the structure of the object and its 3D motion as characterized by rotation and translation. It is this commonality that we wish to exploit by formulating the problems in seemingly disparate applications in a common format and then discussing mathematical tools for solving them.

The list of applications in sensing, modeling, and interpretation of motion and structure from corresponding features observed at different times is rather long. It depends upon the type of features observed. We categorize the problems discussed in this paper and their applications as follows:

- i) 3D-to-3D feature correspondences.
applications:
 - a) Motion estimation using stereo or other range finding devices
 - b) positioning a known 3D object using stereo or other range-finding devices.
- ii) 2D-to-3D feature correspondences.
applications:
 - a) single camera calibration, i.e., determination of position and orientation of a camera knowing some features of 3D objects as imaged on the camera.
 - b) passive navigation of a vehicle using a single camera and based on the knowledge of 3D landmarks.
- iii) 2D-to-2D feature correspondences.
applications:
 - a) finding relative attitudes of two cameras which are both observing the same 3D features.
 - b) Estimating motion and structure of objects moving relative to a camera.
 - c) passive navigation, i.e., finding the relative attitude of a vehicle at two different time instants.

Manuscript received May 23, 1993; revised October 16, 1993.

T. S. Huang is with Beckman Institute and Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801.

A. N. Netravali is with AT&T Bell Laboratories, Murray Hill, NJ 07974.
IEEE Log Number 9214859.

- d) efficient coding and noise reduction of image sequences by estimating motion of objects.

We consider a variety of different features, such as points, straight lines, corners, etc., on the surface of 3D objects. By correspondence is meant observing the same feature at two or many different instants of time or viewing the same feature by two different cameras. In cases where features are 2D, they are obtained usually by perspective transformations of the 3D features on the camera plane. Although most of the paper deals with the perspective transformation, to a lesser extent, orthographic transformation is also considered. For the many cases that result, we formulate the problems, give algorithms for solution, and discuss conditions for existence and uniqueness of solutions and sensitivity of algorithms and solutions to noise. The approach we shall consider for motion/structure determination consists of two steps: i) Extract, match and determine locations (2D or 3D) of corresponding features. ii) Determine motion and structure parameters from the feature correspondences. It is to be emphasized that in this paper we discuss only the second step.

We start with the general statement of the problem and the relevant notation in the next section. We then go through the above three categories of problems, in each case reviewing the formulations and their solutions. The relevant mathematics that deals with solutions of simultaneous linear and nonlinear (especially polynomial) equations in multiple variables is described along the way. Some of these mathematical tools are new to workers in computer vision, but we believe that many of the unsolved questions that we raise in the end require such tools and hope that answers to these questions will be forthcoming using these tools.

II. GENERAL PROBLEM AND NOTATION

Consider an isolated rigid body viewed by an imaging system. Figure 1 shows a typical geometry for imaging. In this figure, object space coordinates are denoted by lower case letters and the image space coordinates are denoted by the upper case letters. Optical center of a pin-hole camera coincides with the origin of a cartesian coordinate system (xyz) and the positive z -axis is the direction of view. The image plane is located at a distance equal to the focal length F (which is assumed to be unity) from o along the direction of view. Using a perspective projection model, a point $p(= (x, y, z))$ on the surface of an object is projected at a point $P(= (X, Y))$ on the image plane, where

$$\begin{aligned} X &= \frac{x}{z} \\ Y &= \frac{y}{z} \end{aligned} \quad (1)$$

In some cases, we also use the orthographic projection instead of the perspective projection. In such cases, the image coordinates of point $P(= (X, Y))$ are related to the

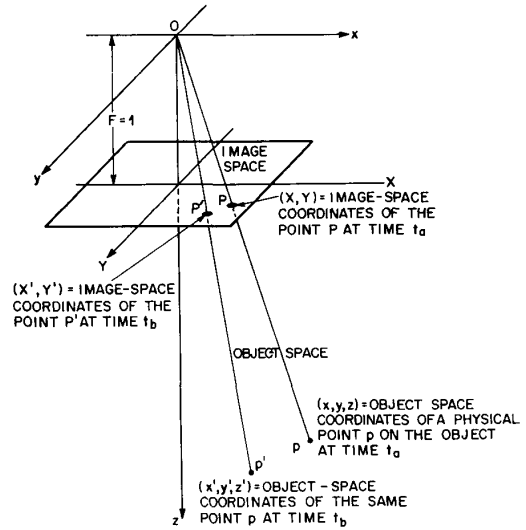


Fig. 1. Basic perspective geometry for imaging. Lower case letters refer to coordinates in the object space and upper case letters refer to coordinates on the image plane. Focal length F is assumed to be equal to one.

object space coordinates of point $p(= (x, y, z))$ by

$$\begin{aligned} X &= x \\ Y &= y. \end{aligned} \quad (2)$$

The problem that we wish to tackle can be stated generally as follows. Consider two features f_1 and f'_1 . These may be points, lines, conics, corners, etc. f_1 and f'_1 refer to description of these features, e.g., their location in either 3 or 2 dimensions. f_1 may be at time t_a and f'_1 is the same corresponding feature at time t_b ; or, f_1 may be specified in 3D space and f'_1 may be its corresponding projection on the image plane. In any case, we are given N such corresponding features on the same rigid object and our problem is to infer motion (where appropriate) and structure of this rigid body with respect to the imaging system. Broadly speaking, there are three different cases:

1) *3D-to-3D Correspondences*: In this case, we are given 3D locations (e.g., object space coordinates in the case of point features) of N corresponding features (points or lines) at two different times and we need to estimate 3D motion of the rigid object. Thus the problem has application in either motion estimation using 3D information which can be obtained by stereo or other range-finding techniques, or positioning a 3D object with known feature locations using stereo techniques.

2) *2D-to-3D Correspondences*: In this case, we are given correspondence of N features (f_i, f'_i) such that f_i are specified in three dimensions and f'_i are their projection on the 2D image plane. The problem is to find location and orientation of the imaging camera. This has applications in either calibration (= determining location and orientation of the camera plane) of a single camera or passive navigation through known 3D landmarks (i.e., f_i) using projection of f_i on a single camera (i.e., f'_i).

3) *2D-to-2D Correspondences*: Here, N corresponding features are specified on the 2D image plane either at two different times for a single camera or at the same instant of time but for two different cameras. In the former case, the problem is to determine 3D motion and structure of the rigid object and in the latter case, the problem is to determine the relative orientation and location of the two imaging cameras.

As mentioned before, in each of the above cases, different formulations are necessary for different kinds of features. In fact, since the problems almost always result in solution of simultaneous linear and nonlinear (mostly polynomial) equations, ability to solve and gain insight depends to a large extent on the proper formulation. Our interest is in pointing out various formulations, algorithms for solution and determining the number of possible solutions as a function of the number of features considered.

A. Motion Model

Since a large part of the paper is concerned with estimating motion of rigid bodies, we state some well known results from kinematics of rigid bodies [3]. Consider a point $p(= (x, y, z))$ at time t_a which moves to location $p'(= (x', y', z'))$ at time t_b as a result of the motion of the rigid body. Then

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathbf{R} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \mathbf{t} \\ = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \quad (3)$$

where \mathbf{R} represents rotation and \mathbf{t} represents translation. Rotation can be specified in a number of equivalent ways. For example, rotation can be around an axis passing through the origin of our coordinate system. Let $\hat{\mathbf{n}} = (n_1, n_2, n_3)^T$ be a unit vector along the axis of rotation and let χ be the angle of rotation from time t_a to time t_b . Then the rotation matrix \mathbf{R} can be expressed as

$$\mathbf{R} = \mathbf{S} + \mathbf{K} \\ \mathbf{S} = \cos \chi \mathbf{I} + (1 - \cos \chi) \hat{\mathbf{n}} \hat{\mathbf{n}}^T \text{ (symmetric part)} \\ \mathbf{K} = \sin \chi \mathbf{N} \text{ (skew-symmetric part)} \quad (4)$$

where \mathbf{I} is the 3×3 identity matrix and

$$\mathbf{N} = \begin{bmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{bmatrix} \quad (5)$$

and $\hat{\mathbf{n}}^T$ is the transpose of $\hat{\mathbf{n}}$.

Alternatively, \mathbf{R} can be specified as three successive rotations around the x -, y -, and z -axis, by angles θ , ψ , and ϕ , respectively, and can be written as a product of these three rotations

$$\mathbf{R} = \begin{bmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \psi & 0 & -\sin \psi \\ 0 & 1 & 0 \\ \sin \psi & 0 & \cos \psi \end{bmatrix} \\ \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix} \quad (6)$$

In the first case, motion is represented by seven parameters $n_1, n_2, n_3, \chi, t_1, t_2$, and t_3 , with a relationship $n_1^2 + n_2^2 + n_3^2 = 1$ since $\hat{\mathbf{n}}$ is a unit vector. In the second case, motion is represented by $\theta, \phi, \psi, t_1, t_2$, and t_3 . Thus in either case, there are six free parameters that express 3D motion. In some cases, we want to specify rotation matrix \mathbf{R} using all its nine components $\{r_{ij}\}_{i,j=1,\dots,3}$. In such a case, it is important to realize that all these nine components are not independent, but the following relationships always hold:

$$\begin{aligned} r_{11}^2 + r_{12}^2 + r_{13}^2 &= 1 \\ r_{21}^2 + r_{22}^2 + r_{23}^2 &= 1 \\ r_{31}^2 + r_{32}^2 + r_{33}^2 &= 1 \\ r_{22}r_{33} - r_{23}r_{32} &= r_{11} \\ r_{23}r_{31} - r_{21}r_{33} &= r_{12} \\ r_{21}r_{32} - r_{22}r_{31} &= r_{13} \end{aligned} \quad (7)$$

Thus in this representation of motion, we have twelve unknowns (nine rotation and three translation), but these six independent relationships among the components of \mathbf{R} result again in six free parameters that express motion.

Finally, we briefly review the quaternion representation of 3D rotation. A rotation around an axis with direction cosines (n_1, n_2, n_3) and rotation angle χ can be represented by the unit quaternion

$$q = (s; l, m, n) = \left[\cos \frac{\chi}{2}; n_1 \sin \frac{\chi}{2}, n_2 \sin \frac{\chi}{2}, n_3 \sin \frac{\chi}{2} \right] \quad (8)$$

specifically

$$(0; p'_i) = q(0; p_i)q^* \quad (9)$$

where $*$ denotes complex conjugation. In terms of q , the rotation matrix becomes (10); see below. We start now with the different cases.

III. 3D-TO-3D CORRESPONDENCES

1) *Point Features*: Our first case is when the features (f_i, f'_i) are 3D coordinates of points on the surface of the rigid body in motion. These are observed at times t_a and

$$\mathbf{R} = \begin{bmatrix} s^2 + l^2 - m^2 - n^2 & 2(lm + sn) & 2(ln + sm) \\ 2(lm + sn) & s^2 - l^2 + m^2 - n^2 & 2(mn - sl) \\ 2(ln - sm) & 2(mn + sl) & s^2 - l^2 - m^2 + n^2 \end{bmatrix} \quad (10)$$

t_b . Suppose we are given N corresponding points (p_i, p'_i) which obey the relationship of (3), i.e.

$$p'_i = R p_i + t, \quad i = 1, \dots, N. \quad (11)$$

The problem is: given (11), find R and t . Since for $N = 1$, the number of independent linear equations is fewer than the number of variables (six), there are always infinitely many solutions. For $N = 2$, there are also infinitely many solutions. It is well known [3] that three noncollinear-point correspondences are necessary and sufficient to determine R and t uniquely.

Equation (11), when expanded represents three scalar equations in six unknown motion parameters. With three point correspondences, we will get nine nonlinear equations. Iterative methods can be used to obtain the "best" fits of the six unknowns. However, it is possible to get stuck in the local minima.

It is therefore advisable to use linear algorithms [4] by observing that (11) is linear in components of R (i.e., r_{ij}) and t . For example, one of the scalar equations from (11) can be written as

$$x'_i = r_{11}x_i + r_{12}y_i + r_{13}z_i + t_1. \quad (12)$$

Therefore, if we have four point correspondences, then we have enough linear equations to solve for the twelve unknowns. In fact, it can be readily seen that if three point correspondences are known, the fourth point correspondence can be found on the basis of rigidity of the body and that if the three points are not collinear then the system of equations (11) is invertible for the unknown parameters.

In practice, since the point correspondences are obtained from measurements, they are subject to error and therefore one prefers to work with more than three point correspondences. In this case, R and t can be obtained as a solution to the following least squares problem:

$$\begin{aligned} &\text{minimize} \left\{ \sum_{i=1}^N \|p'_i - (R p_i + t)\|^2 \right\} \\ &\text{w.r.t. } R, t \end{aligned} \quad (13)$$

subject to the constraint that R is a rotation matrix (e.g., subject to (7)). Here $\|\cdot\|$ represents the Euclidean norm. (As in any such procedure, if there are some measurements with a large amount of error, it is advisable to throw them away, rather than average over them.) Such a constrained least square problem can be solved by linear procedures using quaternions [5], [6] or by singular value decomposition [7], [8].

2) *Motion from Stereo Images Sequences:* Of particular interest is the case of motion estimation from stereo image sequences [10]. Here it is assumed that p_i and p'_i are measured by stereo triangulation. When the ratio of object range (i.e., z) to the camera baseline is large (say > 10), the measurement error due to image sampling is much bigger in z than in x and y [10]. In such a case, weighted least squares can be used to give less importance to z in (13). However, it is difficult to estimate proper weights [11]. Alternatively, maximum likelihood [12] or its approximations [13] can be tried. All these approaches

are computationally time-consuming. It appears that when errors in z are really large, it is better to throw the z 's away and use orthographic approximations [14].

B. 3D-to-3D Straight-Line Correspondences

Suppose, instead of point correspondences, we are given correspondences of straight lines. Thus let l_i and l'_i be two corresponding straight lines at times t_a and t_b , respectively. We note that these lines are considered infinitely long and that no point correspondences on these lines are assumed to be known. Thus our problem is: given corresponding $(l_i, l'_i)_{i=1, \dots, N}$, find R and t .

If $N = 1$, the number of equations is fewer than the number of unknowns and therefore the number of solutions is infinite. Moreover, it can be proved that two nonparallel "sensed" lines are necessary and sufficient to determine R and t uniquely.

The following two types of algorithms may be used.

- i) From the nonparallel sensed lines, we can generate three noncollinear point correspondences and then use the algorithms of Section III-A1.
- ii) Alternatively, if v_i and v'_i are unit vectors along the direction of lines l_i and l'_i ($i = 1, 2$), respectively, and if¹

$$v_3 = v_1 \times v_2 \quad (14)$$

and

$$v'_3 = v'_1 \times v'_2$$

then

$$v'_i = R v_i, \quad i = 1, \dots, 3. \quad (15)$$

This gives us enough equations to determine the rotation matrix R . The translation can be obtained by creating a point correspondence on the lines l_1 and l'_1 . For example, the point on l_1 which is closest to l_2 corresponds to the point on l'_1 which is closest to l'_2 .

To combat noise, rotation can be determined by constrained least squares, i.e.

$$\begin{aligned} &\text{minimize} \left\{ \sum_{i=1}^N \|v'_i - R v_i\|^2 \right\} \\ &\text{w.r.t. } R \end{aligned} \quad (16)$$

subject to R being a rotation matrix. Typically, translation is more sensitive to noise and additional information is necessary to obtain reasonable estimate of the translation [15].

IV. 2D-TO-3D CORRESPONDENCES

As we mentioned in the Introduction, this situation arises when each of the features is 3D and their corresponding features are 2D. If 3D locations of features are known along with their projection on the camera plane which is at unknown location, then the algorithms described in this

¹ \times denotes the usual cross product.

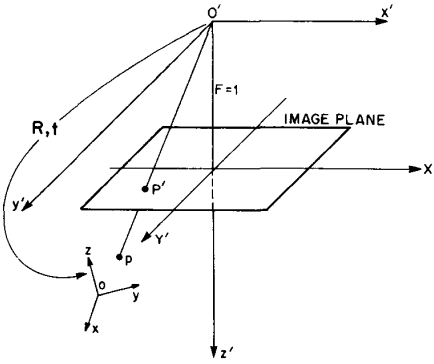


Fig. 2. Imaging geometry for 2D-to-3D correspondence problem. A point p in coordinate system xyz is imaged at location P' on the image plane which is specified in coordinate system $ox'y'z'$.

section allow us to determine the attitude, i.e., the location and the orientation of the camera (known as the camera calibration problem in classical photogrammetry [16]). A related problem is when the camera is attached to a vehicle which is moving in a 3D environment with landmarks at known locations and the goal is to guide this vehicle by observing images of these landmarks. Since the camera is attached to the vehicle, determining the camera attitude, obtains the attitude of the vehicle, which helps navigate it. We consider two types of features: points and straight lines.

A. 2D-to-3D Point Correspondences

As shown in Fig. 2, consider two coordinate systems. xyz is a coordinate system in which the 3D point features are located. Thus p_i are points in this coordinate system with coordinates (x_i, y_i, z_i) . The camera is referenced to the other coordinate system $(o'x'y'z')$. In particular, we assume as before, that the camera plane is perpendicular to the $o'z'$ -axis and at location $z' = 1$. Image coordinates on the camera plane are obtained by perspective projection and denoted by (X', Y') . Thus the image of point p_i is at P'_i whose coordinates are given by

$$X'_i = \frac{x'_i}{z'_i}$$

and

$$Y'_i = \frac{y'_i}{z'_i}. \quad (17)$$

Coordinate system xyz is obtained by a rotation R and translation t of the coordinate system $o'x'y'z'$ and our goal in camera calibration is to determine R and t , knowing the N point correspondences $(p_i, P'_i)_{i=1, \dots, N}$.

The 3D coordinates of p'_i are related to those of p_i by

$$p'_i = R p_i + t \quad (18)$$

or

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R \begin{pmatrix} x \\ y \\ z \end{pmatrix} + t.$$

1) *Church's Method* [16]: Combining (17) and (18), we get

$$\begin{aligned} X'_i &= \frac{r_{11}x_i + r_{12}y_i + r_{13}z_i + t_1}{r_{31}x_i + r_{32}y_i + r_{33}z_i + t_3}, \quad i = 1, \dots, N \\ Y'_i &= \frac{r_{21}x_i + r_{22}y_i + r_{23}z_i + t_2}{r_{31}x_i + r_{32}y_i + r_{33}z_i + t_3}, \quad i = 1, \dots, N. \end{aligned} \quad (19)$$

As in the previous section, there are six unknowns (three for rotation and three for translation) and therefore with three point correspondences, we have enough (i.e., six) equations. Unfortunately, these are nonlinear transcendental equations, since r_{ij} are related to the three unknowns of the rotation matrix (e.g., Euler angles of (6)) in a transcendental manner. Iterative methods are required to solve these nonlinear algebraic equations. In practice, the data (i.e., p_i and P'_i) are known only approximately and therefore one may use more than three point correspondences. In such a case, the following nonlinear least squares problem may be solved iteratively:

$$\begin{aligned} \text{minimize } & \left\{ \sum_{i=1}^N \left(X'_i - \frac{r_{11}x_i + r_{12}y_i + r_{13}z_i + t_1}{r_{31}x_i + r_{32}y_i + r_{33}z_i + t_3} \right)^2 \right. \\ \text{w.r.t. } & R, t \left. \left(Y'_i - \frac{r_{21}x_i + r_{22}y_i + r_{23}z_i + t_2}{r_{31}x_i + r_{32}y_i + r_{33}z_i + t_3} \right)^2 \right\} \end{aligned} \quad (20)$$

subject to R being a rotation matrix (i.e., subject to (7)). The disadvantages of these approaches is that unless one starts with a good initial guess, the iterative procedure may not converge to the right solution. Moreover, no insight is obtained about the nature, uniqueness or the number of solutions.

2) *Ganapathy's Method* [17], [18]: If three noncollinear point correspondences are given, then without loss of generality we can assume that the three points lie in the plane $z = 0$. Then (19) becomes

$$\begin{aligned} X'_i &= \frac{r_{11}x_i + r_{12}y_i + t_1}{r_{31}x_i + r_{32}y_i + t_3} \\ Y'_i &= \frac{r_{21}x_i + r_{22}y_i + t_2}{r_{31}x_i + r_{32}y_i + t_3}. \end{aligned} \quad (21)$$

Thus the three point correspondences give six linear and homogeneous equations in nine unknowns $r_{11}, r_{12}, r_{21}, r_{22}, r_{31}, r_{32}, t_1, t_2, t_3$. Assuming $t_3 \neq 0$, we can divide by t_3 , to get six linear equations in eight unknowns

$$\frac{r_{11}}{t_3}, \frac{r_{12}}{t_3}, \frac{r_{21}}{t_3}, \frac{r_{22}}{t_3}, \frac{r_{31}}{t_3}, \frac{r_{32}}{t_3}, \frac{t_1}{t_3}, \frac{t_2}{t_3}.$$

Additional constraints on $\{r_{ij}\}$ can be obtained from (7) as

$$\begin{aligned} \left(\frac{r_{11}}{t_3} \right)^2 + \left(\frac{r_{21}}{t_3} \right)^2 + \left(\frac{r_{31}}{t_3} \right)^2 \\ = \left(\frac{r_{12}}{t_3} \right)^2 + \left(\frac{r_{22}}{t_3} \right)^2 + \left(\frac{r_{32}}{t_3} \right)^2 \end{aligned} \quad (22)$$

and

$$\left(\frac{r_{11}}{t_3} \right) \left(\frac{r_{12}}{t_3} \right) + \left(\frac{r_{21}}{t_3} \right) \left(\frac{r_{22}}{t_3} \right) + \left(\frac{r_{31}}{t_3} \right) \left(\frac{r_{32}}{t_3} \right) = 0.$$

Thus we have six linear and two quadratic equations in eight unknowns. According to Bézout's theorem [19], the maximum number of solutions (real or complex) is four, assuming that the number of solutions is finite. Solutions can be obtained by computing the resultant which in this case will be a fourth-degree polynomial in one of the unknowns. Note that the unknown "scale factor" t_3 can be determined by using, e.g., the constraint $r_{11}^2 + r_{12}^2 + r_{13}^2 = 1$.

With four coplanar point correspondences, (21) yields eight linear homogeneous equations in nine unknowns. If this system of equations is nonsingular R, t can be obtained uniquely.

If four or five point correspondences are known, then one can either solve a linear least squares problem or use the above method by taking three point correspondences at a time. In the latter case, one expects that the different solution sets have only one common solution. However, precise conditions under which a unique solution is guaranteed are not yet known and in fact, it is easy to construct examples where more than one solution exists with four or five point correspondences.

If six point correspondences are known, then (21) gives twelve linear homogeneous equations in twelve unknowns $\{r_{ij}\}_{i,j=1,\dots,3}$, $\{t_i\}_{i=1,\dots,3}$. R, t can be determined uniquely, if the system is nonsingular. Thus with six or more point correspondences, one expects a unique solution generally.

3) *Fischler and Bolles' Method* [20]: Fischler and Bolles presented a geometric formulation of the problem for three general point correspondences and four coplanar point correspondences. For the three-point case, assume that we are given the three angles $\angle p_2 o' p_3$, and $\angle p_3 o' p_1$ and our problem is to find the lengths $o'p_1$, $o'p_2$, and $o'p_3$. By the law of cosines,

$$\begin{aligned} (\overline{p_1 p_2})^2 &= (\overline{o'p_1})^2 + (\overline{o'p_2})^2 - 2(\overline{o'p_1})(\overline{o'p_2})\cos(\angle p_1 o' p_2) \\ (\overline{p_2 p_3})^2 &= (\overline{o'p_2})^2 + (\overline{o'p_3})^2 - 2(\overline{o'p_2})(\overline{o'p_3})\cos(\angle p_2 o' p_3) \\ (\overline{p_3 p_1})^2 &= (\overline{o'p_3})^2 + (\overline{o'p_1})^2 - 2(\overline{o'p_3})(\overline{o'p_1})\cos(\angle p_3 o' p_1). \end{aligned} \quad (23)$$

Thus we have three quadratic equations in three unknowns, $o'p_1$, $o'p_2$ and $o'p_3$. From Bézout's theorem, the maximum number of solution can be eight. However, because of symmetry, if $(\overline{o'p_1}, \overline{o'p_2}, \overline{o'p_3}) = (a, b, c)$ is a solution, so is $(\overline{o'p_1}, \overline{o'p_2}, \overline{o'p_3}) = (-a, -b, -c)$. Thus there are at most four positive solutions. One can show this, alternatively, by computing the resultant of the system of equations (23) and showing that it is only a fourth-degree polynomial.

In practice, if the measurements are noisy, Fischler and Bolles proposed a robust approach which they called RANSAC.

4) *Horaud, et al.'s Method* [21]: For the four-point case, Horaud, Conio, Le Boulleux, and Lacolle have presented a formulation which results in the solution of a fourth-order polynomial equation in one unknown.

5) *Multiplicity of Solutions and Degeneracy*: Fischler and Bolles [20] showed that with three point correspondences one can have as many as four solutions. Wolfe, Weber-

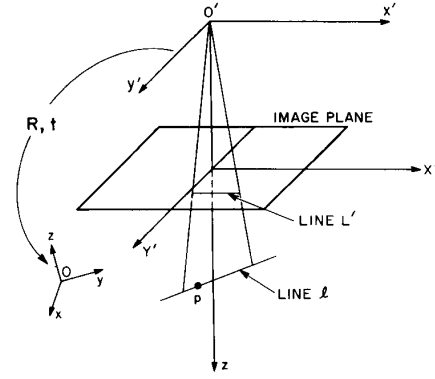


Fig. 3. Imaging geometry for line correspondences. A line l in 3D object space is mapped perspectively into a line L on the 2D image plane.

Sklair, Mathis, and Magee [22] have used a geometrical approach to study this problem of the number of solutions for the three-point case. They exhibited examples of point configurations which give rise to one, two, three, or four solutions. However, to date there is no complete answer to the question: For the N -point case, what geometrical configurations give rise to how many solutions?

Holt and Netravali [23] have given a complete answer to the question: Under what conditions is the number of solutions infinite? The necessary and sufficient condition for infinitely many solutions is essentially that the image points are collinear on the image plane.

B. 2D-to-3D Line Correspondences

Consider the geometry of Fig. 3. Let a straight line l_i in 3D space be perspectively projected on the image plane to give a line L'_i . Also, let the 3D line l_i be characterized by its direction $\mathbf{n}_i = (a_i, b_i, c_i)$ and a point \mathbf{p}_i on it. $(\mathbf{n}_i, \mathbf{p}_i)$ are with respect to the coordinate system xyz . The projected line L'_i is characterized by the following equation in the image plane:

$$A_i X' + B_i Y' + C_i = 0 \quad (24)$$

i.e., by the parameter vector

$$\mathbf{N}'_i \triangleq (A_i, B_i, C_i)^T. \quad (25)$$

As before the two coordinate systems are related by a rotation R and translation t and our goal is to determine them. Thus the problem is as follows: given N correspondences $(l_i, L'_i)_{i=1,\dots,N}$, determine R and t .

1) *Liu, Huang and Faugeras Algorithms* [24]: We first note that the plane containing o' and L' has an equation

$$A_i x' + B_i y' + C_i z' = 0 \quad (26)$$

and therefore the normal of the plane is \mathbf{N}_i . Since l_i lies in this plane, we know that $R\mathbf{n}_i$ is orthogonal to \mathbf{N}'_i and this gives us

$$\mathbf{N}'_i{}^T R\mathbf{n}_i = 0. \quad (27)$$

Thus we have succeeded in decoupling rotation from translation. Since \mathbf{R} contains only three independent variables, with three line correspondences we can solve the nonlinear transcendental equations of (27) iteratively to get \mathbf{R} . Additionally, with eight line correspondences, (27) gives eight linear homogeneous equations in elements of \mathbf{R} (i.e., $\{r_{ij}\}$). If this system of equations is nonsingular, then a unique solution to within a scale factor can be obtained for \mathbf{R} . The scale factor can be determined from the nonlinear constraints on \mathbf{R} of (7). Conditions for nonsingularity are not yet known.

From geometrical considerations, it can be shown that

$$(\mathbf{N}'_i)^T \mathbf{t} = (\mathbf{N}'_i)^T (\mathbf{R} \mathbf{p}_i). \quad (28)$$

Therefore, once \mathbf{R} has been found, (28) can be used to solve for \mathbf{t} from three or more line correspondences.

By decomposing \mathbf{R} in an appropriate way, Chen [25] showed that with three line correspondences, there are at most eight solutions to \mathbf{R} from (27); for each solution of \mathbf{R} , \mathbf{t} can be found from (28). He also demonstrated that eight solutions are attainable for some line configurations. With four, five, six, or seven line correspondences, the exact number of solutions is not yet known. It is expected to be less than eight, since in each of these cases one can take sets of three line correspondences and get eight solutions. The intersection of all these solutions will then be the solution. However, the precise number of these intersections and conditions under which they hold is not yet known. With eight line correspondences, (27) gives a linear system of equations for \mathbf{R} . However, conditions for nonsingularity are not yet known. Chen did succeed in obtaining a necessary and sufficient condition for the number of solutions to be infinite.

More recently, Navab and Faugeras [74] showed that if the feature points lie on a one-sheet hyperboloid passing through the camera center, then the number of solutions to the translation is infinite.

In practice, if the measurements are noisy, by using a larger number of line correspondences, one can pose a variety of least square problems. In most cases, ($3 < N < 8$), the result is a linear least square problem with quadric constraints.

2) *Using Line Correspondence Methods for Point Correspondences*: Given N -point correspondences, we can connect every point pair by a straight line to get $N - (N - 1)/2$ line correspondences, out of which only $3N - 6$ are independent (assuming $N \geq 3$). Then the line correspondence methods of this section can be used to find \mathbf{R}, \mathbf{t} . The advantage of this approach is that \mathbf{R} and \mathbf{t} are decoupled and therefore in a sense the dimensionality of the problem is reduced.

It is interesting to examine the number of solutions obtained by the line and point correspondence method. Three point correspondences generate three line correspondences. However these three lines are coplanar and this appears to ensure that out of the eight possible solutions only four are real. With five point correspondences, we get nine line correspondences and therefore can use the

linear method of line correspondences. Similarly, with six point correspondences, we get twelve line correspondences. Although conditions on uniqueness of solution are not known, empirically a unique solution is obtained for six or more point correspondences.

V. 2D-TO-2D CORRESPONDENCES

Consider a situation where the features of a rigid body are projected on the image plane and only these projections are known. If the camera geometry is fixed, and the object is in motion, then our problem is to obtain 3D motion and structure of this rigid body by observing corresponding projected features at two different instants of time. In this context, by structure is meant the z coordinate (depth or distance along the camera axis) of these features in the 3D space. If, on the other hand, the same rigid body features are observed by projecting them on two different cameras, then the problem is to determine the location and orientation of one camera relative to the other. Both these problems are mathematically equivalent. Different formulations and solutions occur depending on the nature of features. We start with point features.

A. 2D-to-2D Point Correspondences

Consider the problem in which a point $p_i = (x_i, y_i, z_i)$ on a rigid body moves to a point $p'_i = (x'_i, y'_i, z'_i)$ with respect to a camera fixed coordinate system shown in Fig. 4. Let the perspective projection of p_i be $P_i = (X_i, Y_i, 1)$ and that of p'_i be $P'_i = (X'_i, Y'_i, 1)$. Due to the rigid body motion, p_i and p'_i are related by

$$p'_i = \mathbf{R} p_i + \mathbf{t} \quad (29)$$

where \mathbf{R} and \mathbf{t} are rotation and translation, respectively. Our problem is: given N correspondences $(P_i, P'_i)_{i=1, \dots, N}$, determine \mathbf{R} and \mathbf{t} . It is obvious that from two perspective views, it is impossible to determine the magnitude of translation. If the rigid body were two times farther away from the image plane, but twice as big, and translated at twice the speed, we would get exactly the same two images. Therefore, the translation \mathbf{t} and object-point ranges (z_i) can only be determined to within a global positive scale factor. The value of this scale factor can be found if we know the magnitude of \mathbf{t} or the absolute range of any observed object point. Therefore, our problem is to determine rotation \mathbf{R} , $\mathbf{t}/\|\mathbf{t}\|$ and $z_i/\|\mathbf{t}\|$ (= structure). Since the motion and structure are coupled together, different algorithms result depending on whether structure or motion is determined first. We start with the "structure first" algorithm.

1) *"Structure-First" Algorithms* [26]: In this approach, from the rigidity constraints, object-point ranges z_i and z'_i are determined first. Then the motion parameters \mathbf{R} and \mathbf{t} can be determined by solving (29) using the methods of Section III. The points p_i and p'_i are on a rigid body only if the distance between any pair of them does not change in two different views taken at two different times. Thus we must have

$$\|p_i - p_j\|^2 = \|p'_i - p'_j\|^2, \quad 1 \leq i, j \leq N. \quad (30)$$

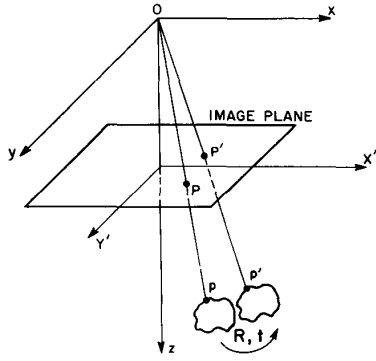


Fig. 4. Imaging geometry for 2D point correspondences. On the image plane, a point P moves to point P' as a result of motion of a rigid body.

For $N(N \geq 3)$ point correspondences, there are $3N - 6$ independent equations. Equation (30) can be expanded as

$$(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 = (x'_i - x'_j)^2 + (y'_i - y'_j)^2 + (z'_i - z'_j)^2. \quad (31)$$

By using the relationships of perspective projection as in (1)

$$(X_i z_i - X_j z_j)^2 + (Y_i z_i - Y_j z_j)^2 + (z_i - z_j)^2 = (X'_i z'_i - X'_j z'_j)^2 + (Y'_i z'_i - Y'_j z'_j)^2 + (z'_i - z'_j)^2. \quad (32)$$

In this equation, unknowns are $\{z_i, z'_i, z_j, z'_j\}$. The equation is a homogeneous quadratic in these unknowns.

With five point correspondences, we have ten unknowns and nine homogeneous equations and iterative algorithms can be used to obtain solutions to within a scale factor. From Bézout's theorem, the number of possible solutions is no more than $2^9 = 512$. Various attempts to obtain bounds on the number of solutions by computing resultants have not been successful. Also, obtaining all the real solutions by using homotopy methods have not been successful due to the large dimensionality of this formulation. For a discussion on resultants, the reader is referred to [27]. For homotopy methods for polynomials, to [28].

2) "Motion-First" Algorithms: In this approach, equations are first formulated that only involve motion parameters. The structure values z_i, z'_i are then obtained from (33) below.

a) Transcendental equations [29], [30]: From (1) and (29), we get (omitting the subscripts)

$$X' = \frac{r_{11}X + r_{12}Y + r_{13} + \frac{t_1}{z}}{r_{31}X + r_{32}Y + r_{33} + \frac{t_3}{z}}$$

and

$$Y' = \frac{r_{21}X + r_{22}Y + r_{23} + \frac{t_2}{z}}{r_{31}X + r_{32}Y + r_{33} + \frac{t_3}{z}}. \quad (33)$$

Eliminating z between these two equations

$$(t_1 - t_3 X')\{(r_{31}X + r_{32}Y + r_{33})Y' - (r_{21}X + r_{22}Y + r_{23})\} = (t_2 - t_3 Y')\{(r_{31}X + r_{32}Y + r_{33})X' - (r_{11}X + r_{12}Y + r_{13})\}. \quad (34)$$

If we use the Euler angle representation of the rotation (6), then (34) is a transcendental equation in terms of these angles and linear and homogeneous in terms of t_i . With five point correspondences, we have five equations in five unknowns (setting the scale factor = 1, e.g., by setting $t_3 = 1$). Iterative methods can be used to solve these nonlinear equations. However, no insight is gained in nature, uniqueness or number of solutions.

b) Polynomial equations: Equation (34) is a second-degree polynomial in eleven unknowns, $\{r_{ij}\}_{i,j=1,\dots,3}, t_1, t_2$ (setting $t_3 = 1$). Then, with five point correspondences, we have five equations from (34) and the six quadratic constraints on the components of the rotation matrix $\{r_j\}$ (7). Thus the number of equations is equal to the number of unknowns and iterative methods can be used for solution. The number of solutions is no more than 2^{11} , but no insight is gained about the number of real solutions. Neither are these particularly suited for numerical computation due to the high dimensionality of search space.

A better approach is due to Jerian and Jain [31] who show that with three point correspondences, a quartic equation can be derived which contains three unknowns; namely, the three components of the "tangent quaternions," representing R . These tangent quaternion components can be expressed as: $n_1 \tan \theta/2$, $n_2 \tan \theta/2$, and $n_3 \tan \theta/2$, where the axis of rotation has direction cosines (n_1, n_2, n_3) and the rotation angle is θ .

Any additional point correspondence gives one more such quartic equation. Therefore, five point correspondences gives three quartic equations in three unknowns. The maximum number of solutions is $4^3 = 64$. In principle, one can reduce the three equations (using resultants) to a single 64th-order equation in one unknown. Then Sturm's method can be used to find all the real solutions. In practice, Jerian and Jain were not able to get the 64th-order equation. They did succeed in reducing two quartics to a single 16th-order equation (using MACSYMA). Thus their method is as follows. Do a global search on one of the three variables. For each fixed value of this variable, the other two variables are solved by reducing two of the three quartics to a single 16th-order equation in one variable whose real roots are determined by Sturm's method. Then each real candidate solution for the three variables is substituted into the remaining quartic equation to check whether it is satisfied. Empirically, Jerian and Jain found that with five point correspondences, the number of real solutions could be two, four, or eight.

c) Linear algorithms [32], [33]: Given eight or more point correspondences, a linear algorithm can be devised. It is obvious from the geometry of Fig. 4, that the three vectors Rp (not shown), p' , and t are coplanar and therefore

$$(p') \cdot (t \times Rp) = 0 \quad (35)$$

which can be written in matrix form as

$$(p')^T E p = 0 \quad (36)$$

where E is a 3×3 matrix defined as

$$E = \begin{bmatrix} e_1 & e_2 & e_3 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{bmatrix} \triangleq GR \quad (37)$$

and G is defined as

$$G \triangleq \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}. \quad (38)$$

Dividing both sides of (36) by the positive quantity zz' (i.e., dividing p by z and p' by z'),

$$(P')^T EP = 0 \quad (39)$$

or

$$[X'Y'1]E \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = 0. \quad (40)$$

Equation (40) is linear and homogeneous in the nine unknowns $\{e_i\}_{i=1,\dots,9}$. Given N point correspondences, we can write (40) in the form

$$B[e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9]^T = 0 \quad (41)$$

where the coefficient matrix B is $N \times 9$. With eight point correspondences, if the rank $(B) = 8$, (41) can be used to uniquely determine E to within a scale factor. Once E is determined, R and t can be determined uniquely, as described, e.g., in [34]. In practice, the data are noisy. Various least squares techniques can be used. See, e.g., [35].

d) *Polynomial methods again:* Given N point correspondences, if the rank $(B) < 8$, then the linear algorithm cannot be used. However, if rank $(B) = 5, 6$, or 7 , then the linear equations (41) can be solved along with the polynomial constraints on the components of matrix E [36], [37]. Specifically, E is equal to a skew-symmetric matrix post multiplied by a rotation matrix only if $\{e_i\}_{i=1,\dots,9}$ satisfy the following three constraint equations. Let $\{\varepsilon_i\}_{i=1,\dots,3}$ be the i th row of E , then

$$\begin{aligned} \varepsilon_3 \cdot (\varepsilon_1 \times \varepsilon_2) &= 0 \\ (||\varepsilon_2||^2 + ||\varepsilon_3||^2 - ||\varepsilon_1||^2)(\varepsilon_2 \cdot \varepsilon_3) + 2(\varepsilon_1 \cdot \varepsilon_2)(\varepsilon_1 \cdot \varepsilon_3) &= 0 \\ ||\varepsilon_3||^4 &= (||\varepsilon_2||^2 - ||\varepsilon_1||^2)^2 + 4(\varepsilon_1 \cdot \varepsilon_2)^2. \end{aligned} \quad (42)$$

Thus we get three polynomial equations in $\{e_i\}_{i=1,\dots,9}$ of degree 3, 4, 4, respectively.

If Rank $(B) = 5$, we have a set of five independent linear homogeneous equations (41) and three homogeneous polynomial equations (42) in the nine unknowns $\{e_i\}$. We can solve these equations to determine $\{e_i\}$ to within a scale factor. From Bézout's theorem, the maximum number of solutions (real or complex) is $3 \times 4 \times 4 = 48$.

One method of solution is to first solve the five linear equations to get (e_1, \dots, e_9) as a linear combination of four known 9-vectors (a basis of the nullspace of B) with unknown weights. Substituting these into (42), we get three homogeneous equations in the four unknown weights. Then,

we solve these three polynomial equations to determine these four weights (to within a scale factor).

If Rank $(B) = 6$, we have a set of six independent linear homogeneous equations in the nine unknowns $\{e_i\}$. We solve these together with the first two equations in (42) to determine $\{e_i\}$ to within a scale factor. The maximum number of solutions (real or complex) is $3 \times 4 = 12$. For each real solution, we substitute it into the third equation of (42) to see whether it is satisfied.

If Rank $(B) = 7$, we have a set of seven independent linear homogeneous equations in the nine unknowns $\{e_i\}$. We solve these together with the first equation in (42). The maximum number of solution is 3. For each real solution we substitute it into the second and third equations in (42) to check if they are satisfied.

For a detailed discussion on such polynomial methods for the cases of Rank $(B) = 5, 6, 7$, see [37].

If Rank $(B) = 8$, then a unique solution to E (to within a scale factor) is ensured. Longuet-Higgins [38] established a necessary and sufficient condition for Rank $(B) = 8$. For convenience, assume the object is stationary and the camera is moving (in particular, the focal point 0 at time t_a moves to $0'$ at time t_b). Then, Rank $(B) = 8$ if and only if the N points lie on a quadratic surface passing through 0 and $0'$. (Note that any nine points lie on a quadratic surface.) Rank $(B) = 8$ is a sufficient but *not* necessary condition for uniqueness of the solution to E .

For Rank $(B) = 5$, which is equivalent to having five point correspondences in general positions, some very loose upper bounds on the number of solutions have been established in earlier sections. Way back in 1913, using projective geometry, Kruppa [39] showed that the number of solutions to E is no more than 11. Recently, Faugeras and Maybank [40], Demazure [41], and Netravali *et al.* [42] sharpened this result to 10.

For Rank $(B) = 6$ or 7 , the number of independent equations in (41) and (42) is larger than the number of unknowns (remembering that the equations are homogeneous so we can set one of the e_i 's equal to 1, thus *heuristically* one would expect that the solution to E is almost always unique. However, it can be shown that for Rank $(B) = 6$ or 7 , it is possible to have up to three solutions in special cases [43], [44].

One point should be clarified here. For a given E matrix (determined to within a scale factor), four pairs of R, t can be derived. However, at most one of the four is physically realizable (i.e., corresponding to 3D points which are in front of the camera both at t_a and t_b) and compatible with the given point correspondences [45].

In practice, since the measurements may be inaccurate and noisy, a larger number of point correspondences should be used. In such cases, a variety of least squares problems can be formulated. Using nonlinear transcendental equations, least squares can be set up as in [46]. If the linear equations (41) are used along with the polynomial constraints (42), then a constrained linear least squares problem has to be solved [47]. Finally, for a large number of points, linear least squares can be formulated [48]. In this

case, error estimates have been derived. One can compute the error estimates alongside the calculation of motion and structure parameters and trust the latter only if the former are sufficiently small.

3) *2D-to-2D Point Correspondences—Planar Case:* Previous sections dealt with situations in which the correspondences of points at any general location were known. If these points have some structure, then this structure can be exploited to generate additional equations and this results in requiring fewer points for solution of the problem. As an example, with four point correspondences, methods of previous sections would yield an infinite number of solutions since the number of equations is smaller than the number of unknowns required to specify motion and structure. However, if the four points are coplanar, then in addition to the six equations from (30), we have the following condition for coplanarity:

$$(\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{p}_1 - \mathbf{p}_3) \times (\mathbf{p}_1 - \mathbf{p}_4) = 0. \quad (43)$$

This gives us seven homogeneous equations in eight unknowns (z_i, z'_i) . From Bézout's theorem, this system can have at most 192 $(= 3 \times 2^6)$ solutions. In fact, as we shall shortly see, the number of solutions is almost always two. In many ways, this conclusion is similar to the 2D-to-3D point correspondence case that we discussed in Section IV-A.

With the motion-first approach, the condition of coplanarity is more difficult to include since perspective projection of coplanar points is not constrained in general on the image plane. Thus with four coplanar points, we still have four linear homogeneous equations from (36). Together with (42), we have seven equations in nine unknowns and therefore if coplanarity is used to generate an additional equation, the number of solutions will be infinite. We do not yet know how to include coplanarity in this formulation. The following linear formulation [49] is more appropriate.

Let the equation of the plane containing the four points at time t_a be given by

$$\mathbf{g}^T \mathbf{p} = 1 \quad (44)$$

where

$$\mathbf{g}^T = (a, b, c). \quad (45)$$

Then at time t_b ,

$$\begin{aligned} \mathbf{p}' &= \mathbf{R}\mathbf{p} + \mathbf{t} \\ &= \mathbf{R}\mathbf{p} + \mathbf{t}\mathbf{g}^T \mathbf{p} \end{aligned} \quad (46)$$

$$\begin{aligned} &= (\mathbf{R} + \mathbf{t}\mathbf{g}^T) \mathbf{p} \\ &= \mathbf{A}\mathbf{p} \end{aligned} \quad (47)$$

where

$$\mathbf{A} = \mathbf{R} + \mathbf{t}\mathbf{g}^T \quad (48)$$

$$\triangleq \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \quad (49)$$

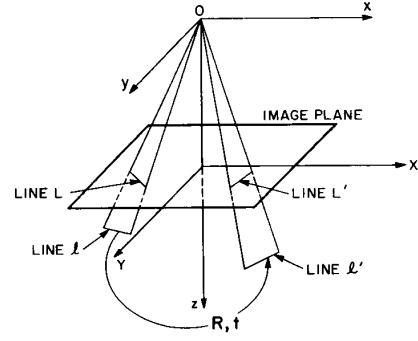


Fig. 5. Imaging geometry for 2D line correspondences. Line l which projects to line L at time t_a moves to line l' at time t_b and projects to line L' .

From (1) and (47)

$$\begin{aligned} X' &= \frac{a_1 X + a_2 Y + a_3}{a_7 X + a_8 Y + a_9} \\ Y' &= \frac{a_4 X + a_5 Y + a_6}{a_7 X + a_8 Y + a_9}. \end{aligned} \quad (50)$$

Thus with four point correspondences, we have eight homogeneous linear equations in nine unknowns $\{a_i\}_{i=1,\dots,9}$ from (50). It can be easily shown that if no three points out of the four are collinear, then this system of linear equations has rank 4. Therefore, we have a unique solution to $\{a_i\}$ to within a scale factor. Once matrix \mathbf{A} is determined to within a scale factor, $\{\mathbf{R}, \mathbf{w}\mathbf{t}, \mathbf{g}/\mathbf{w}\}$ can be determined, where \mathbf{w} is an unknown positive scale factor [49]. It is shown in [49], [50] that there are in general only two solutions. It is also easy to see that additional coplanar points will not yield any more linearly independent equations in $\{a_i\}$. Also, it is known that if we track four coplanar points over three views, then the solution is unique [51]. In practical situations, where the point correspondences may be known only approximately, a least squares algorithm as described in [52] may be used, where a procedure for computing the error estimates is also given.

B. 2D-to-2D Line Correspondences

Consider the geometry of Fig. 5, where l_i is the 3D line on a rigid object at time t_a and L_i is its perspective projection on the image plane. As a result of object motion, this 3D line l_i becomes line l'_i in three dimensions at time t_b and its corresponding projection on the image plane becomes L'_i . Our problem then is as follows: given $\{L_i, L'_i\}_{i=1,\dots,N}$, the corresponding lines on the image plane, determine the rotation \mathbf{R} , translation $\mathbf{t}/\|\mathbf{t}\|$, and the 3D line positions to within a scale factor, e.g., $\|\mathbf{t}\|$.

We shall see presently that the number of solutions is infinite no matter how large N is. One way to make solutions unique (or finite in number) is to take three views, i.e., take the projection at another time t_c . Thus let l''_i and L''_i be the corresponding lines at t_c . Assume that the motion from t_a to t_b and t_b to t_c is represented by $\{\mathbf{R}_{ab}, \mathbf{t}_{ab}\}$ and $\{\mathbf{R}_{bc}, \mathbf{t}_{bc}\}$, respectively. Our problem

is then modified to: given three corresponding 2D views $\{L_i, L'_i, L''_i\}_{i=1,\dots,N}$ of lines, determine motion parameters $\{R_{ab}, R_{bc}, \kappa t_{ab}, \kappa t_{bc}\}$ and the 3D positions at t_a of the lines l_i , where the κ is a scale factor. As before, we assume that no point correspondences on the lines are known.

1) *Structure-First Algorithm* [53]: Let us consider the two-view case first. At each time instant t_a and t_b , given the projected line L , the 3D line l is restricted to lie in the plane containing the origin O and line L . Two additional parameters are needed to fix l . Thus there are four "structure" parameter unknowns (two at t_a and t_b , each) associated with each line. Given two lines, we can write two equations representing the rigidity constraints which imply that the angle and the distance between the two lines remains constant from time t_a to t_b . Any additional line adds four more unknowns and four more rigidity equations (representing the angle and distance constraints between the new line and two of the old lines). Thus no matter how many line correspondences we have, the number of equations is always smaller (by 2) than the number of unknowns. One would expect then that the number of solutions is infinite.

Now consider the three-view case (t_a, t_b , and t_c). Given N line correspondences over three views, the number of "structure" unknowns is $6N$. The number of equations is $2 \times 2 + (N - 2) \times 8 = 8N - 12$. The number of equations is larger than or equal to the number of unknowns when $N \geq 6$. Thus at least six line correspondences over three views are needed to get a finite number of solutions to the motion/structure parameters. Heuristically, seven or more line correspondences over three views probably make the solution unique (to within a scale factor for the translation and structure parameters), although no proof is yet available.

2) *A Motion-First Algorithm* [54], [55]: Consider the geometry of Fig. 6, where 3D lines l, l' , and l'' (at times t_a, t_b , and t_c , respectively) are projected on the image plane to get 2D lines L, L' and L'' . Let M, M' , and M'' be the planes containing origin O and lines L, L' , and L'' , respectively. If the lines L, L' , and L'' are represented by

$$\begin{aligned} L : AX + BY + C &= 0 \\ L' : A'X + B'Y + C' &= 0 \\ L'' : A''X + B''Y + C'' &= 0 \end{aligned} \quad (51)$$

then the normals to planes M, M' , and M'' are given by $N^T = (A, B, C)$, $N'^T = (A', B', C')$, and $N''^T = (A'', B'', C'')$, respectively. Obviously, N, N' , and N'' are perpendicular to l, l', l'' , respectively. If we rotate M (together with l) by R_{ab} and rotate M'' (together with l'') by R_{bc}^{-1} , then $R_{ab}l, l'$ and $R_{bc}^{-1}l''$ will be parallel to each other, which implies that $R_{ab}N, N'$ and $R_{bc}^{-1}N''$ are coplanar, i.e.,

$$(R_{ab}N) \cdot N' \times (R_{bc}^{-1}N'') = 0. \quad (52)$$

This is a nonlinear (transcendental) equation in the six rotation unknowns (e.g., the Euler angles). With six line correspondences over three views, we can solve the resulting six equations (52) iteratively to find the rotations.

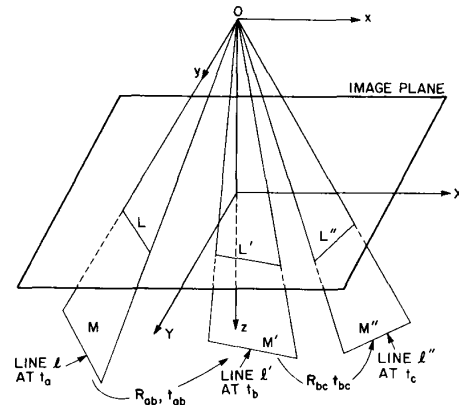


Fig. 6. Line correspondences for the three-view system.

It can be shown [55] that the translations satisfy

$$t_{ab} \cdot (R_{ab}N) = \frac{\|N' \times R_{ab}N\|}{\|N' \times R_{bc}^{-1}N''\|} R_{bc}^{-1}t_{bc} \cdot (R_{bc}^{-1}N'') \quad (53)$$

which is a linear homogeneous equation in the six translation components (after the rotations have been found). Thus the translations can be determined to within a scale factor from five line correspondences over three views.

We note that the problem of finding the rotations can be formulated as a solution of polynomials in several variables. Equation (52) is quadratic (and homogeneous) in the 18 unknowns which are the elements of R_{ab} and R_{bc}^{-1} . With six line correspondences over three views, we have six equations similar to (52) plus the 12 constraint equations on the elements of the rotation matrices. Thus we have a total of 18 quadratic equations in 18 unknowns. This admits a maximum of 2^{18} solutions, by Bezout's Theorem. However, only a few of these will be real, but this number is not yet known. An upper bound of 600 was established recently by Holt and Netravali [75].

Buchanan [56], [76] has established some results on the conditions under which the line case fails to get a unique solution. Specifically, he showed that the algorithm based on (52) will fail to get a unique solution if the lines lie on a line complex, and that any algorithm will fail if the lines lie on a line congruence. The rather abstract results of Buchanan have been given some nice geometrical interpretation by Navab *et al.* [77].

3) *A Linear Algorithm* [57], [58]: For convenience, we shall try to determine (R_{ab}, t_{ab}) and (R_{ac}, t_{ac}) , the latter being the motion from t_a to t_c . We use $R^{(i)}$ to denote the i th column of R . Let

$$F_i \triangleq t_{ac}(R_{ab}^{(i)})^T - (R_{ac}^{(i)})^T t_{ab}, \quad i = 1, 2, 3. \quad (54)$$

Then it can be shown that

$$\begin{aligned} AN'' \cdot F_2N' - BN'' \cdot F_1N' &= 0 \\ BN'' \cdot F_3N' - CN'' \cdot F_2N' &= 0 \\ CN'' \cdot F_1N' - AN'' \cdot F_3N' &= 0 \end{aligned} \quad (55)$$

where A, B, C are as defined in (51). The three equations in (55) are linear and homogeneous in the $3 \times 9 = 27$ elements of \mathbf{F}_i ; $i = 1, 2, 3$. However, only two of the equations are linearly independent. Thus we need 13 line correspondences over three views to get 26 equations (55) to determine the elements of \mathbf{F}_i to within a scale factor. After \mathbf{F}_i have been found (to within a scale factor), the motion parameters can be obtained via (54) as described in [59].

It is shown in [59] that the linear algorithm gives a unique solution except in certain degenerate cases. However, a geometrical interpretation of the degenerate cases is not yet available. In practical situations, due to noisy data, robust numerical algorithms are required. The polynomial approach leads to the problem of finding least squares solutions to a set of quadratic equations subject to quadratic equality constraints. Sufficient numerical experience with such problems is not yet available. Linear algorithms can be made robust more easily as shown in [59].

C. 2D-to-2D Point Correspondences: Orthographic Projections

If a point $\mathbf{p} = (x, y, z)$ in 3D space is projected orthographically to a point $\mathbf{P} = (x, y)$ on the image plane, then

$$X = x$$

and

$$Y = y. \quad (56)$$

The problem then is as follows: given corresponding points $\{\mathbf{P}_i, \mathbf{P}'_i\}_{i=1, \dots, N}$, determine rotation \mathbf{R} , translation \mathbf{t} , and depth z_i . A little reflection indicates that t_3 is irrecoverable and the z_i 's can be recovered at best to within an additive constant.

We can decompose the motion into a rotation \mathbf{R} around one of the points (say \mathbf{p}_1) and translation $\mathbf{t} = \mathbf{p}'_1 - \mathbf{p}_1$. Then the x - and y - components of \mathbf{t} can be found simply from $\mathbf{p}'_1 - \mathbf{p}_1$. Thus without loss of generality, we can fix \mathbf{p}_1 and \mathbf{p}'_1 at the origin and the problem is then reduced to finding \mathbf{R} and $\{z_i\}_{i=2, \dots, N}$. Therefore, our modified problem is as follows: given correspondences $\{\mathbf{P}_i, \mathbf{P}'_i\}_{i=2, \dots, N}$ and assuming $\mathbf{P}_1^T = \mathbf{P}_0^T = (0, 0, 0)$, determine \mathbf{R} and $\{z_i\}_{i=2, \dots, N}$.

1) *Two-View Case [60]*: Since $\mathbf{p}'_i = \mathbf{R}\mathbf{p}_i$, $i = 2, \dots, N$ and from (56), we have

$$\begin{aligned} X'_i &= r_{11}X_i + r_{12}Y_i + r_{13}z_i \\ Y'_i &= r_{21}X_i + r_{22}Y_i + r_{23}z_i. \end{aligned} \quad (57)$$

Eliminating z_i and using properties of \mathbf{R} , we get

$$r_{23}X'_i - r_{13}Y'_i + r_{32}X_i - r_{31}Y_i = 0 \quad (58)$$

which is linear and homogeneous in the four unknowns: $r_{23}, r_{13}, r_{32}, r_{31}$. With four point correspondences, we have three eqations (58). Assuming the four points are non-coplanar, it can be shown that the coefficient matrix has rank 3. Thus we can determine $(r_{13}, r_{23}, r_{31}, r_{32})$ uniquely

to within a scale factor. Note that any additional point correspondences will only give redundant (58).

By changing the scale factor, we can get an uncountably infinite number of solutions to $(r_{13}, r_{23}, r_{31}, r_{32})$ which satisfy

$$r_{13}^2 + r_{23}^2 = r_{31}^2 + r_{32}^2 < 1 \quad (59)$$

and for each of these solutions we can construct a rotation matrix \mathbf{R} . Thus our conclusion is that with two orthographic views, there are an uncountably infinite number of solutions to motion/structure, no matter how many point correspondences are given.

2) *Three-View Case*: Ullman [61] showed that four point correspondences over three orthographic views yield a unique solution to motion/structure (up to a reflection). He also provided a nonlinear algorithm for finding the solution. More recently, Lee [62] and Huang and Lee [60] have found a number of linear algorithms. The basic idea of the linear algorithm in [60] is the following: Let

$$\begin{aligned} \mathbf{P}' &= \mathbf{R}\mathbf{p} \\ \mathbf{P}'' &= \mathbf{S}\mathbf{p}' \end{aligned} \quad (60)$$

and

$$\mathbf{P}'' = \mathbf{W}\mathbf{p}$$

where

$$\mathbf{W} = \mathbf{S}\mathbf{R}. \quad (61)$$

Then, taking two views at a time and using the method of Section V-C1, we can find $(r_{13}, r_{23}, r_{31}, r_{32})$, $(s_{13}, s_{23}, s_{31}, s_{32})$, and $(w_{13}, w_{23}, w_{31}, w_{32})$ to within scale factors. Then these three scale factors are determined from the constraint equation (61). In [60], [63], it is also shown that three point correspondences over three orthographic views yield four solutions to the structure and 16 solutions to the motion parameters (including reflections).

3) Relationship to Perspective Projections:

- i) As mentioned in Section III, in determining motion/structure from stereo image sequences (perspective projection model), the ranges of the points (obtained from triangulation) can be very inaccurate. Thus it may be a good idea to disregard z 's (at least initially), and use orthographic-view techniques [14].
- ii) In Section V-A we have seen that with five or more point correspondences over two *perspective* views, motion/structure can be determined. On the other hand, we have shown in the previous section that motion/structure cannot be determined no matter how many point correspondences we have over two orthographic views. Since orthographic projection is a reasonable approximation to perspective projection when the object is relatively far away from the camera, we conclude that when the object is relatively far away from the camera, the motion/structure from point correspondences over two perspective views must be ill-conditioned.

- iii) A thorough analysis is needed to find out when and how the perspective projections can be approximated by orthographic projections (perhaps with a scale change). In this regard, [64] is relevant.

D. 2D-to-2D Correspondences: Other Features

1) *Combined use of Points and Lines* [65]: When both point and line correspondences are given over two views, the method of using rigidity constraints described in Sections V-A and V-B can be readily extended. The additional rigidity constraints between points and lines can, for example, be written in terms of the invariance of the distance from a point to a line.

It appears that adding line correspondences to point correspondences will not help (in the absence of noise), since each line brings in four additional unknowns and four additional equations. That each line yields four additional equations can be seen as follows. On the rigid configuration of N points ($N \geq 3$), we can establish a 3D coordinate system. To specify a 3D line in this coordinate system, we need four parameters.

In the presence of noise, least squares techniques are typically used. Then, the addition of line correspondences may help in getting more accurate estimates of motion/structure.

2) *Using Corners and Curves*: In [66], orthogonal corners are used as features. Generally, one orthogonal corner and two point or line correspondences determine motion/structure uniquely. Another feature used in [67], [68], [78] is conic arcs. The value of these variety of features is yet unknown both in terms of our ability to recognize them and then track them.

VI. FUTURE RESEARCH AND OPEN QUESTIONS

A. Noise Sensitivity

In all practical situations, the coordinates of features cannot be measured exactly. The errors in feature coordinates cause errors in motion/structure estimation. It is well known that the noise sensitivity problem can be especially severe in the 2D-to-2D feature correspondence case. However, good results can be obtained in many situations [35].

Ideally, one would like to have analytical results (both algorithm-dependent and algorithm-independent) on how noise sensitivity depends on feature configurations and motion characteristics. Unfortunately, no such results are available. Some heuristic and empirical results do exist [34], [35]. For example, it is reported that the estimates are very sensitive to data noise, if the object occupies a small part of the field of view, or if the z -component of the translation is small.

For linear algorithms, one can perform a numerical first-order error analysis along with the motion determination [35], [52]. Thus one obtains not only the motion/structure parameters but also estimates of the errors contained in them. Only if the error estimates are small, will one accept the motion/structure results. This approach becomes

difficult, when the motion/structure determination algorithm is nonlinear.

B. Robust Algorithms

There are two general approaches to robustness: Least squares and RANSAC. In least squares, one tries to smooth out noise by using as much data as possible. However, in motion/structure determination, some of the feature correspondences may be incorrect. These "outliers" should be weeded out by, for example, using least squares with adaptive weights. RANSAC [20] takes the approach of using as few data as possible to obtain a good solution. To illustrate it, take the case of motion/structure estimation using 2D-to-2D point correspondences. Assume 20 point correspondences are given. In RANSAC, one would pick six point correspondences and determine the motion parameters. Then, these computed motion parameters will be checked against the remaining 14 point correspondences for consistency. If they are consistent with the majority, one would accept them. If not, one picks a different set of six point correspondences, and so on, until one gets a solution which is consistent with a majority of the 20 given point correspondences.

We have seen from the previous sections that problems of motion/structure determination lead to the solution of simultaneous equations. These equations can be transcendental (e.g., if rotation is represented by Euler angles) or polynomial (e.g., if rotation is represented by a quaternion). Even in the absence of noise, transcendental equations can be solved only by iterative methods. For polynomial equations (when the number of equations is equal to the number of unknowns), one has the option of using homotopy methods [28]. When the data are noisy (which is always the case in practice), we face the problem of nonlinear least squares. In principle, a polynomial nonlinear least squares problem can be reduced to the solution of a larger set (than the original set of polynomial equations) of polynomial equations which can be solved by homotopy methods. However, in reality, even in the simplest motion/structure estimation problems, the total degree of this new polynomial equation set will be too large to handle numerically. Thus the only option appears to be to solve the nonlinear least squares problem by iterative methods. A major difficulty is of course that unless one has a good initial guess at the solution, the iteration may lead to a local but not global minimum, or may not converge at all.

An alternative formulation of motion/structure determination problems leads to linear least squares with polynomial (usually quadratic) equality constraints (e.g., see the end of section V-A2d [47].) Another example is presented in (21) and (22) in Section IV-A2. Unfortunately, no particularly efficient algorithm for solving such problems is known.

As we have seen in the previous sections, in a number of cases, it is possible to formulate linear algorithms at the expense of a small number of additional feature correspondences. Then the least squares problems become linear. The computation is simpler, and above all one

does not have to worry about being trapped in local minima. However, it has been found empirically that linear algorithms are usually more sensitive to data noise than nonlinear algorithms.

On balance, perhaps the best approach is to first use a linear algorithm (assuming that a sufficient number of feature correspondences are given) to find a rough solution; and then to use a nonlinear formulation to refine the solution iteratively [46].

C. Multiplicity of Solutions and Degeneracy

In problems of motion/structure determination, it is of both theoretical and practical interest to know how the number of solutions depends on the number of feature correspondences. Let N be the number of feature correspondences, and $m(N)$ be the number of solutions at a “generic” data set. Then, typically, for a particular problem, there is a number N_o such that: For $N < N_o$, $m(N) = \infty$ (uncountable). For $N = N_o$, $m(N) \leq m_o = \text{finite positive integer}$. For $N > N_o$, $m(N) = 1$. We have assumed that the problem arises from a real physical situation so that there is at least one real solution; we have also assumed that there is no noise in the data. The critical case $N = N_o$ is of particular interest, because: 1) In practice, it is not easy to obtain feature correspondences. Thus it is important to know the minimum number needed. 2) To use RANSAC-like techniques, one needs to know N_o .

Two important related questions are: a) Nonuniqueness-degeneracy: For $N > N_o$, what are the necessary and sufficient conditions (on the feature configurations and the motion characteristics) such that $1 < m(N) < \infty$? b) ∞ -degeneracy: For $N \geq N_o$, what are the necessary and sufficient conditions such that $m(N) = \infty$? These questions are closely related to noise sensitivity (Section VI-A). Specifically, for $N > N_o$, if the feature configuration and motion characteristics are close to be degenerate, then the estimation problem will most likely be ill-conditioned.

Unfortunately, answers to these important questions are far from complete. Our present knowledge is summarized in Table 1. In addition to the question marks in the table, it is also not known, for the 2D-to-3D and 2D-to-2D cases, under what conditions, $m(N_o) = 1, 2, \dots, m_o$, respectively.

The mathematical tools appropriate for attacking these questions appear to be algebraic geometry [27], [42] and projective geometry [40], [69]. Numerically, homotopy methods seem to be the best for finding all solutions of a set of simultaneous polynomial equations [28], and “upper semicontinuity” arguments [63], [79] have the potential of proving general theorems by exhibiting numerical examples.

D. Long Sequences

In this paper, we have concentrated on motion/structure determination using only two (or sometimes three) time instants or frames. This is sufficient for some applications

Table 1 Solution Multiplicity and Degeneracy Conditions

	N_o	$m(N_o)$	Conditions for nonuniqueness-degeneracy	Conditions for ∞ -degeneracy
3D to 3D PC	3	1	collinear	collinear
3D to 3D LC	2	1	parallel	parallel
2D to 3D PC	3	≤ 4 [20]	?	Essentially collinear in 2D [23]
2D to 3D LC	3	≤ 8 [24]	?	Some parallelism conditions [25]
2D to 2D PC	5	≤ 10 [40],[42]	3D points lying on certain quadratic surface [43],[44]	?
2D to 2D LC (over 3 frames)	6	?	?(some partial results in [56])	?

Note: References are in square brackets.

(e.g., passive navigation, pose determination, camera calibration), but not for others (e.g., motion prediction). For motion prediction and general understanding, it is necessary to work with longer image sequences. Furthermore, using longer image sequences is potentially a way of combating noise in the data.

The key issue in the analysis of long image sequences is motion modeling. For a given scenario, one needs a motion model containing a small number of parameters which can be assumed to remain constant during a short period of time. Significant work in this area has been reported in [70]–[73]. Here, we shall discuss only the work of Shariat [73].

Shariat used a “constant object motion” model, where the object is assumed to be translating with a constant velocity T and rotating with a constant R around an unknown object center $C = (x_c, y_c, z_c)$. The parameters to determine are R, T , and C . He derived the following main results: 1) One point over five frames: seven second-order homogeneous polynomial equations in eight unknowns. 2) Two points over four frames: seven second-order homogeneous polynomial equations in eight unknowns. 3) Three points over three frames: eight second-order homogeneous polynomial equations in nine unknowns. Two important open questions are: i) Multiplicity of solutions and degeneracy conditions? ii) Effective linear algorithms?

VII. SUMMARY AND CONCLUSION

We have presented a review of techniques for determining motion and structure of rigid bodies, by knowing the locations of their corresponding features at different times

or when they are projected on two different cameras. Three major categories of problems were considered: a) 3D features are known at two different times; b) features in 3D space and their projections on the camera plane are known; c) projections of 3D features on the camera plane are known at two different times. Features may consist of points, straight lines, curved lines, corners, etc. In each case, we reviewed a variety of formulations, efficient algorithms for solution, existence and uniqueness of solutions, and sensitivity of solutions to noise in the data. We showed that, even though problems and their formulation may be very different, for each of the above cases, the underlying mathematics has much in common. It requires solution of simultaneous, transcendental, polynomial, or linear equations in multiple variables which represent the structure, and motion of the object. Thus the problems are inherently non-linear and therefore appropriate formulation is extremely important to avoid difficulties in either numerical computation of the solution or determination of nonuniqueness and multiplicity of solutions. A variety of theoretical (e.g., algebraic and projective geometry) and numerical (e.g., homotopy) techniques are being applied to these problems with varying degrees of success. Despite over two decades of work in modern time and half century of work around 1900 (by photogrammetrists) problems are far from being completely solved. Section VI contains a partial list of some of the open questions. Techniques reviewed in this paper have found a number of applications. Some of these are: a) positioning and navigating objects in a 3D world; b) Determining location and orientation of a camera by observing 3D features in its view; c) Estimating motion and structure of moving objects relative to a camera; and there are many more. We believe that the new mathematical tools will allow us to make steady progress in the future towards the resolution of many of the outstanding questions and this technology will get applied in a wide variety of situations.

REFERENCES

- [1] T. S. Huang, "Motion analysis," in *Encyclopedia of Artificial Intelligence*, S. Shapiro, Ed. New York: Wiley, 1987.
- [2] J. K. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images—A review," *Proc. IEEE*, vol. 76, no. 8, pp. 917–935, Aug. 1988.
- [3] H. Goldstein, *Classical Mechanics*. Reading, MA: Addison Wesley, 1981.
- [4] S. D. Blostein and T. S. Huang, "Algorithms for motion estimation based on 3-D correspondences," in *Motion Understanding*, W. Martin and J. K. Aggarwal, Eds. Norwell, MA: Kluwer, 1988.
- [5] O. D. Faugeras and M. Hebert, "A 3-D recognition and positioning algorithm using geometrical matching between primitive surfaces," in *Proc. Int. Joint Conf. on Artificial Intelligence* (Karlsruhe, W. Germany), Aug. 1983, pp. 996–1002.
- [6] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Amer.*, vol. 4, pp. 629–642, Apr. 1987.
- [7] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point set," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, no. 5, pp. 698–700, Sept. 1987.
- [8] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," to appear in *IEEE Trans. Pattern Anal. Machine Intell.*
- [9] T. S. Huang *et al.*, "Motion/structure estimation from stereo image sequences," in *Proc. IEEE Workshop on Motion* (Kiawah Island, SC, May 7–9), 1986.
- [10] S. D. Blostein and T. S. Huang, "Quantization errors in stereo triangulation," in *Proc. 1st ICCV* (London, England, June 8–11, 1987), pp. 325–344. Also, in *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 752–765, 1987.
- [11] L. Matthies and S. Shafer, "Error modeling in stereo navigation," *IEEE Trans. Robotics Automat.*, vol. 3, pp. 239–248, 1987.
- [12] C. W. Chen and T. S. Huang, "Error probability based estimation of motion from stereo sequences," in *Proc. 3rd Int. Workshop on Time-varying Image Processing* (Florence, Italy), May 1989.
- [13] S. M. Kiang, R. J. Chou, and J. K. Aggarwal, "Triangulation errors in stereo algorithms," in *Proc. IEEE Workshop on Computer Vision* (Miami Beach, FL, Dec. 1987), pp. 72–78.
- [14] T. S. Huang, "Use of orthographic-projection techniques in motion estimation from stereo image sequences," *Machine Vision for Inspection and Measurement*, H. Freeman, Ed. New York: Academic Press, 1989, pp. 127–135.
- [15] H. H. Chen and T. S. Huang, "An algorithm for matching 3-D line segments with applications to multiple object motion estimation," in *Proc. IEEE Workshop on Computer Vision* (Miami Beach, FL, Nov. 30–Dec. 2, 1987), pp. 151–156.
- [16] P. R. Wolf, *Elements of Photogrammetry*. New York: McGraw-Hill, 1974.
- [17] S. Ganapathy, "Camera location determination problem," Tech. Memo., AT&T Bell Laboratories, Holmdel, NJ, Nov. 2, 1984.
- [18] S. Ganapathy, "Decomposition of transformation matrices for robot vision," *Pattern Recogn. Lett.*, vol. 2, pp. 401–412, Dec. 1989.
- [19] B. L. van der Waerden, *Modern Algebra*, vol. 2. Berlin, Germany: Springer, 1940.
- [20] M. Fischler and "R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, June 1981.
- [21] R. Horaud, B. Conio, O. Le Boulleux, and B. LaColle, "An analytic solution for the perspective 4-point problem," *Comput. Vis., Graphics, Image Process.*, vol. 47, no. 1, pp. 33–44, July 1989.
- [22] W. J. Wolfe, C. Weber-Sklair, D. Mathis, and M. Magee, "The perspective view of three points," to appear in *IEEE Trans. Pattern Anal. Machine Intell.*
- [23] R. J. Holt and A. N. Netravali, "Camera calibration problem: Some new results," Tech. Memo., AT&T Bell Laboratories, Murray Hill, NJ, July 1989.
- [24] Y. Liu, T. S. Huang, and O. D. Faugeras, "Determination of camera location from 2D to 3D line and point correspondences," in *Proc. CVPR*, (Ann Arbor, MI, June 1988).
- [25] H. H. Chen, "Locating polyhedral objects with dextrous hand by straight finger motion," Tech. Memo., AT&T Bell Laboratories, Holmdel, NJ, May 1989.
- [26] A. Mitiche and J. K. Aggarwal, "A computational analysis of time-varying images," in *Handbook of Pattern Recognition and Image Processing*, T. Y. Young and K. S. Fu, Eds. New York: Academic Press, 1986.
- [27] W. V. D. Hodge and D. Pedoe, *Methods of Algebraic Geometry*, 3 vols. London, UK: Cambridge Univ. Press, 1953.
- [28] A. Morgan, *Solving Polynomial Systems using Continuation for Engineering and Science Problems*, Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [29] J. W. Roach and J. K. Aggarwal, "Determining the movement of objects from a sequence of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 2, no. 6, pp. 554–562, 1980.
- [30] T. S. Huang and R. Y. Tsai, "Image sequence analysis: Motion estimation," in *Image Sequence Analysis*, T. S. Huang, Ed. New York: Springer-Verlag, 1981.
- [31] C. Jerian and R. Jain, "Polynomial methods for structure from motion," in *Proc. 2nd ICCV* (Tampa, FL), Dec. 1988.
- [32] H. C. Longuet-Higgins, "A computer program for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133–135, Sept. 1981.
- [33] 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. 6, no. 1, pp. 13–27, Jan. 1984.

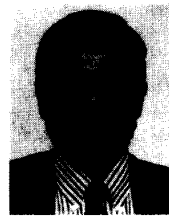
- [34] T. S. Huang, "Determining 3 dimensional motion and structure from two perspective views," in *Handbook of Pattern Recognition and Image Processing*, T. Y. Young and K. S. Fu, Eds. New York: Academic Press, 1986.
- [35] J. Weng, T. S. Huang, and N. Ahuja, "Motion and structure from two perspective views: algorithms, error analysis, and error estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 5, pp. 451-476, May 1989.
- [36] T. S. Huang and Y. S. Shim, "Linear algorithm for motion estimation: How to handle degenerate cases," in *Proc. British Pattern Recognition Association Conf.* (Cambridge, England), Apr. 1987. Also published by Springer-Verlag, J. Kittler, Ed.
- [37] T. S. Huang and O. D. Faugeras, "Some properties of the e-matrix in two-view motion estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 12, pp. 1310-1312, Dec. 1989.
- [38] H. C. Longuet-Higgins, "The reconstruction of a scene from two projections-configurations that defeat the 8-point algorithm," in *Proc. 1st Conf. on AI Applications* (Denver, CO, Dec. 5-7), 1984, pp. 395-397.
- [39] E. Kruppa, "Zur Ermittlung eines Objektes Aulzwei Zwei Perspektieren Mit Innerer Orientierung," *Sitz-Ber. Akad. Wiss., Wien, math. naturw. kl., Abt-IIa*, vol. 122, pp. 1939-1948, 1913.
- [40] O. D. Faugeras and S. J. Maybank, "Multiplicity of solutions for motion problems," Tech. Rep., INRIA, France, 1988. Appeared later in *Proc. 1st ICCV* (Irvine, CA, Mar. 20-22), 1989, pp. 248-255.
- [41] M. Demazure, "Sur Deux Problèmes de Reconstruction," Tech. Rep., INRIA, France, 1988.
- [42] A. N. Netravali, T. S. Huang, A. S. Krishnakumar, and R. J. Holt, "Algebraic methods in 3D motion estimation from two-view point correspondences," *Int. J. Imaging Syst. Technol.*, vol. 1, pp. 78-99, 1989.
- [43] B. K. P. Horn, "Motion fields are hardly ever ambiguous," *Int. J. of Comput. Vision*, vol. 1, pp. 263-278, 1987.
- [44] H. C. Longuet-Higgins, "Multiple interpretation of a pair of images of a surface," *Proc. Royal Soc. London, Series A*, vol. 418, pp. 1-15, 1988.
- [45] M. A. Snyder, "On the calculation of rigid motion parameters from the essential matrix," COINS Tech. Rep. TR 89-102, Univ. of Massachusetts at Amherst, Jan. 1990.
- [46] J. Weng, N. Ahuja, and T. S. Huang, "Closed-form solution and maximum likelihood: A robust approach to motion and structure estimation," in *Proc. IEEE Conf. on CVPR* (Ann Arbor, MI, June 5-9), 1988.
- [47] C. Braccini, G. Gambardella, A. Grattarola, and S. Zappature, "Motion estimation of rigid bodies: effects of the rigidity constraints," in *Proc. EUSIPCO-86* (The Hague, The Netherlands, Sept. 2-5), 1986.
- [48] J. Weng, N. Ahuja, and T. S. Huang, "Error analysis of motion parameter estimation from image sequences," in *Proc. 1st Int. Conf. on Computer Vision* (London, England), June 1987.
- [49] R. Y. Tsai, T. S. Huang, and W. L. Zhu, "Estimating 3-D motion parameters of a rigid planar patch 2, II: Singular value decomposition," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 30, no. 4, pp. 523-525, Aug. 1982.
- [50] H. C. Longuet-Higgins, "The visual ambiguity of a moving plane," *Proc. Roy. Soc., Ser. B*, vol. 223, no. 1, pp. 165-170, 1984.
- [51] R. Y. Tsai and T. S. Huang, "Estimating 3-D motion parameters of a rigid planar patch, III: Finite point correspondences and the three-view problem," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, no. 2, pp. 213-220, Apr. 1984.
- [52] J. Weng, N. Ahuja, and T. S. Huang, "Motion and structure from point correspondences: a robust algorithm for planar case with error estimation," in *Proc. 9th ICPR* (Rome, Italy, Nov. 14-17), 1988.
- [53] A. Mitiche and J. K. Aggarwal, "Line-based computation of structure and motion using angular invariance," in *Proc. IEEE Workshop on Motion* (Kiawah Island, SC, May 7-9), 1986, pp. 175-180.
- [54] B. L. Yen and T. S. Huang, "Determining 3-D motion and structure of a rigid object using straight line correspondences," in *Image Sequence Processing and Dynamic Scene Analysis*, T. S. Huang, Ed. New York: Springer-Verlag, 1983.
- [55] Y. C. Liu and T. S. Huang, "Estimation of rigid body motion using straight line correspondences," in *Proc. IEEE Workshop on Motion* (Kiawah Island, SC, May 7-9), 1986, pp. 47-52.
- [56] T. Buchanan, "Notes on two recent reconstruction algorithms," unpublished report, 1987.
- [57] M. E. Spetsakis and J. Aloimonos, "Closed form solutions to the structure from motion problem from line correspondences," in *Proc. AAAI* (Seattle, WA, July 1987), pp. 738-743.
- [58] Y. Liu and T. S. Huang, "A linear algorithm for motion estimation using straight line correspondences," in *Proc. 9th ICPR* (Rome, Italy, Nov. 14-17), 1988.
- [59] J. Weng, Y. Liu, T. S. Huang, and N. Ahuja, "Estimating motion/structure from line correspondences: A robust algorithm and uniqueness theorem," in *Proc. IEEE Conf. on CVPR* (Ann Arbor, MI, June 5-9, 1988).
- [60] T. S. Huang and C. H. Lee, "Motion and structure from orthographic projections," in *Proc. 9th ICPR* (Rome, Italy, Nov. 14-17), 1988. A longer version appeared in *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 5, pp. 536-540, May 1989.
- [61] S. Ullman, *The Interpretation of Visual Motion*. Cambridge, MAAA: MIT Press, 1979.
- [62] C. H. Lee, "Structure from motion: An augmented problem and a new algorithm," Tech. Rep., Dept. of CS, Purdue Univ., Sept. 1, 1986.
- [63] D. D. Hoffman and B. M. Bennett, "The construction of structure from fixed-axis motion: rigid structures," *Biol. Cybern.*, vol. 54, pp. 71-83, 1986.
- [64] J. Aloimonos and M. Swain, "Paraperspective projection: Between orthograph and perspective," Tech. Rep. CAR-TR-320, Center for Automation Research, Univ. of Maryland, College Park, MD, May 1987.
- [65] J. K. Aggarwal and Y. F. Wang, "Analysis of a sequence of images using point and line correspondences," in *Proc. Int. Conf. on Robotics and Automation* (Raleigh, NC, Mar. 1987).
- [66] Y. Liu and T. S. Huang, "Motion estimation from corner correspondences," in *Proc. Int. Conf. on Image Processing* (Singapore, Sept. 5-8, 1989), pp. 785-790.
- [67] R. Y. Tsai, "Estimating 3-D motion parameters and object surface structure from the image motion of curved edges," in *Proc. IEEE Conf. on CVPR* (Washington, DC), 1983, pp. 259-266.
- [68] R. Y. Tsai, "3-D inference from the motion parallax of a conic arc and a point in two perspective views," in *Proc. 8th IJCAI* (Karlsruhe, W. Germany), pp. 1039-1042.
- [69] J. G. Semple and G. T. Kneebone, *Algebraic Projective Geometry*. Oxford, England: Clarendon Press, 1952.
- [70] A. N. Netravali and J. Salz, "Algorithms for estimation of three-dimensional motion," *AT&T Tech. J.*, pp. 335-346, Feb. 1985.
- [71] T. S. Huang, J. Weng, and N. Ahuja, "3D motion from image sequences: Modeling, understanding, and prediction," in *Proc. IEEE Workshop on Motion* (Kiawah Island, SC, May 7-9), 1986, pp. 125-130.
- [72] T. J. Broida and R. Chellappa, "Kinematics and structure of a rigid object from a sequence of noisy images," in *Proc. IEEE Workshop on Motion* (Kiawah Island, SC, May 7-9), 1986, pp. 95-100.
- [73] H. Shariat, "The motion problem: How to use more than two-frames," Rep. IRIS-202, Univ. of Southern California, Oct. 6, 1986.
- [74] N. Navab and O. D. Faugeras, "Monocular pose determination from lines: critical sets and maximum number of solutions," in *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (New York, June 15-17, 1993), pp. 254-260.
- [75] R. J. Holt and A. N. Netravali, "Motion and structure from line correspondences: some further results," AT&T Bell Labs. Tech. Memo., Sept. 29, 1992.
- [76] T. Buchanan, "Critical sets for 3D reconstruction using lines," in *Proc. 2nd European Conf. on Computer Vision* (Santa Margherita, Italy, May 1992), pp. 730-738.
- [77] N. Navab, "Visual motion of lines, and cooperation between motion and stereo," Dissertation, Univ. of Paris XI, Orsay, France, Jan. 1993.
- [78] Song De Ma, "Conic based stereo, motion estimation, and pose determination," preprint, May 1992.
- [79] R. J. Holt and A. N. Netravali, "Motion from optic flow: Multiplicity of solutions," *J. Visual Commun. Image Represent.*, vol. 4, no. 1, pp. 14-24, Mar. 1993.



Thomas S. Huang (Fellow, IEEE) 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 MIT 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, West Lafayette, IN, 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, and at the Beckman Institute. During his sabbatical leaves he has worked at the MIT Lincoln Laboratory, Lexington, MA, the IBM T.J. Watson Research Center, Yorktown Heights, NY, and the Rheinisches Landes Museum in Bonn, West Germany, and held visiting Professor positions at the Swiss Institutes of Technology in Zurich and Lausanne, University of Hannover, in West Germany, INRS-Telecommunications of the University of Quebec in Montreal, Canada, and the University of Tokyo, Japan. 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 11 books and over 300 papers in Network Theory, Digital Filtering, Image Processing, and Computer Vision. He has received a Guggenheim Fellowship (1971–1972), an A. V. Humboldt Foundation Senior U.S. Scientist Award (1976–1977), and a Fellowship from the Japan Association for the Promotion of Science (1986). He received the IEEE Signal Processing Society's Technical Achievement Award in 1987 and the Society Award in 1991. He was a founding Editor, and is currently an Area Editor of the *International Journal of Computer Vision, Graphics, and Image Processing*, and Editor of the Springer Series in Information Sciences published by Springer-Verlag.

Dr. Huang is a Fellow of the Optical Society of America.



Arun Netravali (Fellow, IEEE) received the B.Tech. (Honors) degree from the Indian Institute of Technology, Bombay, India, in 1967, and the M.S. and Ph.D. degrees from Rice University, Houston, TX, in 1969 and 1970, respectively, all in electrical engineering.

He was at NASA from 1970 to 1972, where he worked on problems related to filtering, guidance, and control for the Space Shuttle. He joined Bell Laboratories in 1972 as a Member of Technical Staff, became Head of the Visual Communications Research Department in 1978, Director of Computing Systems Research in 1983, and assumed his current position as Executive Director of Research Communications Sciences Division and Technology Conversion Laboratories at AT&T Bell Laboratories, Murray Hill, NJ, in 1992 with added responsibility as a project manager for HDTV since 1990. His responsibility is research in all aspects of communication and networking. He has been an adjunct professor at the Massachusetts Institute of Technology since 1984, and has taught graduate courses at City College, New York, Columbia University, New York, MIT, Cambridge, MA, and Rutgers University, Piscataway, NJ. He served on the Editorial Board of the *PROCEEDINGS OF THE IEEE* from 1980 to 1984, and is currently an editor of several journals. He is an advisor to the Center for Telecommunications Research of Columbia University, the Swiss Federal Institute of Technology in Lausanne, Switzerland, and the Beckman Institute of the University of Illinois at Urbana-Champaign. He has organized and chaired sessions at several technical conferences and was the Program Chairman of the 1981 Picture Coding Symposium and the 1990 International Conference on Pattern Recognition. He has edited several special issues for IEEE publications, including two for *PROCEEDINGS OF THE IEEE* (Digital Encoding of Graphics, Visual Communication Systems) and one for *IEEE TRANSACTIONS ON* (Picture Communication Systems). He is the author of more than 100 papers and holds over 50 patents in the areas of Computer Networks, Human Interfaces to Machines, Picture Processing, and Digital Television. He is the co-author of two books: *Digital Picture Representation and Compression* (New York, Plenum, 1987) and *Visual Communication Systems* (New York, IEEE PRESS, 1989). He received the Donald G. Fink Award for the best review paper published in *PROCEEDINGS OF THE IEEE* in 1980, the journal award for the best paper from the Society of Motion Pictures and Television Engineers in 1982, the L. G. Abraham Award for the best paper from the IEEE Communications Society in 1985 and 1991, the Alexander Graham Bell Medal in 1991, the OCA National Corporate Employee Achievement Award in 1991, and Engineer of the Year Award from the Association of Engineers from India in 1992. He serves on the New Jersey Governor's Committee on "Schools" programs.

Dr. Netravali is a member of Tau Beta Pi and Sigma Xi. He is also a Fellow of the AAAS, and a member of the United States National Academy of Engineering. He has served on the Digital Television Committees of the IEEE and the Society of Motion Picture and Television Engineers.