1988

Structure and Motion from Two Perspective Views Via Planar Patch

Chia-Hoang Lee

Report Number:

88-734

STRUCTURE AND MOTION FROM TWO PERSPECTIVE VIEWS VIA PLANAR PATCH

Chia-Hoang Lee

CSD-TR-734
January 1988

STRUCTURE AND MOTION FROM TWO PERSPECTIVE VIEWS VIA PLANAR PATCH

Chia-Hoang Lee
Department of Computer Science
Purdue University
West Lafayette, IN 47907

Abstract

We present a method which uses both a planar patch and points for an analysis of time-varying imagery. From our theory, two pieces of information are required for the unique recovery of structure from motion. One is the image motion of a plane and the other one is the image motion of points not on the plane. For the former, one needs four points while for the latter one need two points. Our technique has the following characteristics: (i) it is efficient, (ii) it appears to be very robust over a wide range of simulations, (iii) it sheds some new light on some previous work, and (iv) it has geometric meaning for each step of the derivation of the solution.

1. INTRODUCTION

Deriving 3D motion parameters and the structure of an object from time-varying imagery is an important task in computer vision. On one hand, it has a broad range of practical applications such as target tracking, autonomous vehicle navigation, and robot guidance. On the other hand, it facilitates the understanding of and suggests possible hypotheses for human motion perception. As a result, this line of study has received extensive attention from researchers.

To solve such a problem, two basic approaches are frequently used: methods based on optical flow, and methods based on line or point matches. The former is called the flow-based approach while the latter is called the feature-based approach. This classification is by no means exhaustive, some other approaches use range sensors or structure light. In the feature-based approach, each frame of the sequence is segmented first, and the feature points/lines resulting from the projection of vertices and edges of the object surfaces are marked. Next, the correspondence of these features between the successive frames is established. Lastly, the motion parameters and object structure are derived. See [1,4,9,10,11] for further references on the feature-based method. For the flow-based method, see [5,7,12].

One of the important problems for the recovery of structure from motion is to develop reliable methods. In this regard, Ullman [11] suggests integrating information over longer time intervals than the minimum necessary for a unique solution. Another possibility is to integrate a different class of features.

Recently researchers have begun to make use of a multitude of features on the object for computing the structure and motion of 3D objects. In particular, Aggarwal and Wang [1] studied the use of both lines and points for 3D motion analysis. They argue that there is no compelling reason to prefer any single class of features and ignore others. This line of research is interesting because none of the existing methods based on a single class of features has been shown to be robust.

We shall present a method which uses both a planar patch and points for an analysis of time-varying imagery. Our technique has the following characteristics: (i) it is efficient, (ii) it appears to be very robust over a wide range of simulations, (iii) it sheds some new light on some previous work, and (iv) it has geometric meaning for each step of the derivation of the solution.

2. PROBLEM

We assume that the image plane is stationary and that two perspective views, at times t_1 and t_2, are taken of a rigid object moving in the 3-D object space. The object has two parts: a planar patch Ω which consists of four points and two other points which do not lie on the plane defined by Ω. When there is no possibility of confusion, we shall also use Ω to denote the plane defined by Ω. We shall further assume that all six points are in view. Our purpose is to derive the motion and structure of the 3-D object from the two views.

We shall use the following notation. The focal length f will be assumed to be 1, which implies that the image plane is at a distance 1 along the positive z axis from a camera origin. Let

$z_i A_i$ = Object-space coordinates of a point P_i on the rigid object at t_1

$z'_i B_i$ = Object-space coordinates of the same point P_i at t_2

A_i = Image-space coordinates of the point P_i at t_1

B_i = Image-space coordinates of the point P_i at t_2

Notice that the third component of A_i and B_i is 1. Then

$$z'_i B_i = R z_i A_i + T \quad i=1,..,6$$

where

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$ is a rotation matrix,

and

$$T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$ is a translation vector,

It is easy to see that if (R, T, z_i, z'_i) is a solution then (R, cT, cz_i, cz'_i) is also a solution where c is a scalar. Thus one could at best derive z_i, z'_i and T up to a scale. With this, we shall normalize T such that $t_z = 1$ if $t_z \neq 0$. Of course, whether $t_z = 0$ is still unkonwn.

The problem we are trying to solve is: Given 6 image point correspondences

$$A_i <\!\!\rightarrow B_i \quad ; \qquad i = 1, 2, ..., 6$$

where $z_i\ A_i$, $i = 1, 2, 3, 4$ are coplanar, determine R, T, and z_i, z'_i, $i = 1, \cdots 6$. Figure 1 depicts the imaging geometry and the problem. We also assume that no three of A_i's or B_i's are collinear for $i = 1,2,3,4$. In other words, the camera (origin) is not on the plane defined by the planar patch. Thus the planar patch should form a quadrilateral in the image plane.

3. FEATURE-BASED METHOD: PREVIOUS WORK

The techniques based on orthographic projections will not be discussed here for two reasons: (1) our technique uses a perspective projection model, (2) the set of solutions based on two views using these two techniques are quite different. With orthographic projection, [13] infinitely many solutions always exist for any number of point correspondences, while finite or unique solutions can be found for a finite number of point correspondences in the case of perspective projection.

Examing approaches in the literature, one could classify them into three categories in terms of input formulation. The first one uses point (vertex) correspondences, the second one uses line (edge) correspondences, and the third one uses both point and line correspondences.

Among these three, methods based on point correspondences are more closely related to our formulation in which an additional planar requirement is imposed over several points. It is also interesting to note that this planar assumption frequently causes degenerate configurations that defeat existing algorithm [3].

Roach and Aggarwal [6] show that five points in two views are needed to recover the structure and motion parameters. Their approach requires solving a system of 18 nonlinear equations with 27 variables. This method requires iterative search and a good initial guess of a solution. Nagel and Neuman [4] observe that the rotation matrix can be separated from the translation matrix. The idea stems from the observation that $Rz_iA_i \times z'_iB_i$, if not zero, defines a vector normal to a plane containing T ,where $\times$ stands for vector product and i=1,..5. From this, we conclude that $(RA_1 \times B_1) \times (RA_2 \times B_2)$ is oriented in the same direction as T. Therefore $(RA_1 \times B_1) \times (RA_2 \times B_2) \cdot (RA_k \times B_k) = 0$ for k=3,4,5, which is a set of fourth-order polynomial equations in three unknown (parameters of rotation). This technique requires many fewer search dimensions than that in [6].

Tsai and Huang [9] have proposed a method to find the motion of a planar patch (containing at least four points) from 2-D perspective views. The technique consists of two steps: First, a set of eight "pure parameters" is defined. These parameters can be determined uniquely from two successive image frames by solving a set of linear equations. Then, the actual motion parameters are determined from these eight "pure parameters" by solving a sixth-order polynomial. The authors reported that aside from a scale factor for the translation parameters, the number of real solutions never exceeded two on their simulations.

The problem of a curved surface patch in motion was investigated in [10]. Here two main results of the method referred to as the 8-points algorithm were established concerning the existence and uniqueness of the solutions. First, given the image correspondences of eight object points in general position, an E matrix can be determined by solving eight linear equations. Furthermore, the actual 3-D motion parameters can be determined uniquely from E. In other words, as long as E is unique, the 3-D motion parameters are unique. A similar algorithm (not addressing the aspect of uniqueness) was also discovered independently by Longuest-Higgins [2]. Higgins [3] furthermore enumerates inherent configurations that lead to singularities of E, hence 3D motion parameters is not necessarily unique. For instance: if any six of the points lie on a conic, if any four of the points are collinear, if any seven of the points lie in a plane. Tsai and Huang [10] also includes following conclusion: The image correspondences of four points on a plane not passing through the origin and two other points not on this plane determine the motion parameters uniquely.

In veiw of the previous approaches, the following strategy could solve our problem. First, apply the technique in [9] to the planar patch to obtain solutions. Second, choose the solutions that conform to the other two points. The uniqueness of the solution is then ensured by the 8-point algorithm. However, the method in [9] requires very high accuracy in the input, which often makes it impractical.

4. METHODS

Our results consist of three steps. In the first step, we derive a technique which could compute the motion on the image for every point on the plane defined by Ω. Recall that A_5, A_6 are projections of points not on Ω. However, we could consider two virtual points on Ω with A_5 and A_6 as projections. Next we use the technique to predict the projections of these two virtual points on the image at time t_2. The second step uses these two predictions and B_5, B_6 to derive the translational parameters. The third step constructs a rotation matrix from both the derived translation and the technique developed in the first step. The depth z_i, z'_i thus follows.

4.1. Image Motion of the Plane Ω: An Algebraic Approach

Let A, B be the projections of the same point in the plane defined by Ω at time t_1 and t_2. We present a technique to compute B from A using four pairs of markings A_i and B_i ($i = 1, \cdots 4$) that are projections of points in Ω.

Recall that

$$R\ z_1 A_1 = z'_1 B_1 - T \quad (1)$$
$$R\ z_2 A_2 = z'_2 B_2 - T \quad (2)$$
$$R\ z_3 A_3 = z'_3 B_3 - T \quad (3)$$
$$R\ z_4 A_4 = z'_4 B_4 - T \quad (4)$$

Since A and B are projections of the same point in Ω at two instants, there exist z and z' such that zA and $z'B$, respectively, lie on Ω at t_1 and t_2. The relation between zA and $z'B$ is governed by the following equation:

$$R\, z\, A = z' B - T \tag{5}$$

Our current goal is to derive B from A_i, B_i $(i = 1, \cdots 4)$ and A without knowing R, T, z_i, z'_i $(i = 1, \cdots 4)$, z, z'.

Since $z_4 A_4$ is coplanar with $z_1 A_1, z_2 A_2, z_3 A_3$, we know that there exist a_1, a_2, a_3 such that

$$a_1 + a_2 + a_3 = 1$$

and

$$a_1 z_1 A_1 + a_2 z_2 A_2 + a_3 z_3 A_3 = z_4 A_4 . \tag{6}$$

Applying R to both sides of the above equation and using (1)-(4), we rewrite it as

$$a_1 z'_1 B_1 + a_2 z'_2 B_2 + a_3 z'_3 B_3 = z'_4 B_4. \tag{7}$$

Dividing both sides of (6) and (7) by z_4 and z'_4 respectively, one gets

$$a_1 \frac{z_1}{z_4} A_1 + a_2 \frac{z_2}{z_4} A_2 + a_3 \frac{z_3}{z_4} A_3 = A_4$$

and

$$a_1 \frac{z'_1}{z'_4} B_1 + a_2 \frac{z'_2}{z'_4} B_2 + a_3 \frac{z'_3}{z'_4} B_3 = B_4.$$

Let $A_{ij} = A_i \times A_j$; $B_{ij} = B_i \times B_j$ where $\times$ denotes vector product, $i, j = 1, \cdots 3$ and $i < j$.

It is easy to see that

$$a_k \frac{z_k}{z_4} A_k \cdot A_{ij} = A_4 \cdot A_{ij} \tag{8}$$

$$a_k \frac{z'_k}{z'_4} B_k \cdot B_{ij} = B_4 \cdot B_{ij} \tag{9}$$

where $k \neq i$; $k \neq j$; $k = 1, 2, 3$; and $i < j$.

Since no three of A_1,A_2,A_3, and A_4 are collinear, we divide (9) by (8) and obtain

$$\frac{z'_k}{z_k} = \frac{A_k \cdot A_{ij}}{B_k \cdot B_{ij}} \frac{B_4 \cdot B_{ij}}{A_4 \cdot A_{ij}} \frac{z'_4}{z_4}$$

for $k = 1, 2, 3$ and $k \neq i, j$.

Following simple calculations, one gets

$$\frac{z'_2}{z_2} = \frac{A_2 \cdot A_{13}}{B_2 \cdot B_{13}} \cdot \frac{B_4 \cdot B_{13}}{A_4 \cdot A_{13}} \cdot \frac{B_1 \cdot B_{23}}{A_1 \cdot A_{23}} \cdot \frac{A_4 \cdot A_{23}}{B_4 \cdot B_{23}} \frac{z'_1}{z_1} \overset{def}{\equiv} \delta_2 \frac{z'_1}{z_1}$$

$$\frac{z'_3}{z_3} = \frac{A_3 \cdot A_{12}}{B_3 \cdot B_{12}} \cdot \frac{B_4 \cdot B_{12}}{A_4 \cdot A_{12}} \cdot \frac{B_1 \cdot B_{23}}{A_1 \cdot A_{23}} \cdot \frac{A_4 \cdot A_{23}}{B_4 \cdot B_{23}} \frac{z'_1}{F_1} \overset{def}{\equiv} \delta_3 \frac{z'_1}{z_1}.$$

We now compute B from A. Since $z\ A$ is coplanar with $z_1 A_1$, $z_2 A_2$, $z_3 A_3$, there exists b_1, b_2, b_3 and $b_1 + b_2 + b_3 = 1$ such that

$$b_1 z_1 A_1 + b_2 z_2 A_2 + b_3 z_3 A_3 = z\ A \tag{10}$$

Dividing both sides of (10) by z, we obtain

$$b_1 \frac{z_1}{z} A_1 + b_2 \frac{z_2}{z} A_2 + b_3 \frac{z_3}{z} A_3 = A.$$

Therefore,

$$b_k \frac{z_k}{z} = \frac{A \cdot A_{ij}}{A_k \cdot A_{ij}} \quad i, j, k = 1, 2, 3 \text{ and } k \neq i, j\ ; \quad i < j.$$

Applying R to (10), we get

$$b_1 z'_1 B_1 + b_2 z'_2 B_2 + b_3 z'_3 B_3 = z' B \quad .$$

Rewriting it into

$$b_1 \frac{z'_1}{z_1} \frac{z_1}{z} B_1 + b_2 \frac{z'_2}{z_2} \frac{z_2}{z} B_2 + b_3 \frac{z'_3}{z_3} \cdot \frac{z_3}{z} B_3 = \frac{z'}{z} B.$$

and factoring $\frac{z'_1}{z_1}$ out, we obtain

$$\frac{z'_1}{z_1} \left(b_1 \frac{z_1}{z} B_1 + b_2 \frac{z_2}{z} \delta_2 B_2 + b_3 \frac{z_3}{z} \delta_3 B_3 \right) = \frac{z'}{z} B.$$

The left hand side, apart from a scale $\frac{z'_1}{z_1}$, may now be computed because all the coefficients of the B_i's are known. Furthermore the third component of B (right hand side) must be 1. Hence B may be derived.

We will denote B by $M_\Omega(A)$. This notation means that (i) A is an image point; (ii) if we regard A as the projection of a virtual point on Ω, then $M_\Omega(A)$ is the projection of the same virtual point at time t_2. If the roles of these two views are reversed, we can obtain $M^{-1}{}_\Omega(\cdot)$ in the same way.

4.2. Image Motion of the Plane Ω: A Geometric Drawing

In this subsection, we show that we can also use geometric drawing to determine image motion of the plane Ω. Consider Figure 2, which contains two views of a planar patch. The problem is to predict where A is in the second view given that A is the projection of a point on Ω. First, one finds that P, the intersection of lines A_3A_1 and A_2A_4, is mapped to Q, the intersection of lines B_3B_1 and B_2B_4. Second, one draws the line joining A and A_3, which intersects lines A_2A_4 and A_2A_1 at U and V respectively. Since A_2A_4 and B_2B_4 are two projections of the same segment, the invariance of the cross ratio can be applied here. This says that the cross ratio defined by A_4,P,U, and A_2 is equal to the cross ratio defined by B_4,Q,S, and B_2, where S is yet to be found. The following relation describes the equation:

$$\frac{A_4A_2 \cdot PU}{A_4U \cdot PA_2} = \frac{B_4B_2 \cdot QS}{B_4S \cdot QB_2}$$

Once S is determined, one draws the line joining B_3 and S, which intersects B_1B_2 at T. It is clear that the line A_3U is mapped to the line B_3T. Now one has A_3,U,V, and A on one projection and B_3,S,and T on the other projection, so B could be determined by using the invariance of the cross ratio again.

4.3. Deriving Translational Parameters

In this section, we will present three theorems which show how to derive the translational parameters based on $M_\Omega(\cdot)$ discussed above. Recall that $M_\Omega(\cdot)$ is determined by A_i and B_i ($i = 1, \cdots 4$); A_5 corresponds to B_5; A_6 corresponds to B_6. Also A_5 and A_6 are projections of points not on Ω. These facts will be assumed in the following theorems.

Theorem 1:

If $T = (a, b, 0)^t$ then $T \,//\, M_\Omega(A_5) - B_5$ and $T \,//\, M_\Omega(A_6) - B_6$, where // denotes "is parallel to" and $a^2 + b^2 \neq 0$.

Proof:

Recall that $R\ z_5 A_5 = z'_5 B_5 - T$. Since the tip of the vector $z_5 A_5$ emanating from the camera origin does not lie on Ω, there exists a scalar $c \neq 1$ such that $c\ z_5 A_5$ lies on Ω. Multiplying both sides of the above equation by c, one gets

$$R\ c\ z_5 A_5 = c\ z'_5 B_5 - c\ T\ = (\ c\ z'_5 B_5 - cT + T\ ) - T.$$

It is clear that the third component of $c\ z'_5 B_5 - c\ T + T$ is $c\ z'_5$ and the above equation can be written as

$$R\ c\ z_5 A_5 = c\ z'_5 \left[\frac{c\ z'_5 B_5 - c\ T + T}{c\ z'_5} \right] - T.$$

Thus

$$M_\Omega\ (A_5) = \frac{c\ z'_5 B_5 - c\ T + T}{c\ z'_5}.$$

Next one examines $M_\Omega\ (A_5) - B_5$:

$$M_\Omega\ (A_5) - B_5 = \frac{c\ z'_5 B_5 - c\ T + T}{c\ z'_5} - B_5 = \frac{(1-c)}{c\ z'_5}\ T,$$

which gives $M_\Omega\ (A_5) - B_5 \ // \ T$. By the same reasoning, we can show that $M_\Omega\ (A_6) - B_6 \ // \ T$. Q.E.D.

Theorem 2:

$M_\Omega\ (A_5) = B_5$ is a necessary and sufficient condition for $T = (0, 0, 0)^t$.

Proof:

From theorem 1, we have

$$M_{\Omega}(A_5) - B_5 = \frac{c\ z'_5 B_5 - c\ T + T}{c\ z'_5} - B_5 = \frac{(1-c)}{c\ z'_5}\ T.$$

Thus $M_{\Omega}(A_5) = B_5$ is a necessary and sufficient condition for $T = (0, 0, 0)'$ Q.E.D.

Theorem 3:

If $T = (a, b, 1)'$, then T (regarded as a point in the image plane) lies on the line joining $M_{\Omega}(A_5)$ and B_5. T also lies on the line joining $M_{\Omega}(A_6)$ and B_6.

Proof:

Repeat the argument used in the proof of Theorem 1 up to

$$R\ c\ z_5 A_5 = (\ c\ z'_5 B_5 - cT + T\ ) - T.$$

The third component of $c\ z'_5 B_5 - c\ T + T$ is $c\ z'_5 - c + 1$, thus this equation can be written as

$$R\ c\ z_5 A_5 = (\ c\ z'_5 - c + 1\ )\ \frac{(c\ z'_5 B_5 - c\ T + T\ )}{c\ z'_5 - c + 1} - T.$$

It is clear that

$$M_{\Omega}(A_5) = \frac{c\ z'_5 B_5 - c\ T + T}{c\ z'_5 - c + 1}\quad .$$

Examining $M_{\Omega}(A_5) - B_5$:

$$M_\Omega(A_5) - B_5 = \frac{c\,z'_5 B_5 - c\,T + T}{c\,z'_5 - c + 1} - B_5 = \frac{c-1}{c\,z'_5 - c + 1}(B_5 - T) \quad ,$$

we thus conclude that T must lie on the line joining $M_\Omega(A_5)$ and B_5. The same reasoning also applies to A_6 and B_6 Q.E.D.

It is interesting to see that there are two ways of writing $M_\Omega(A)$: one allows the computation of $M_\Omega(A)$ while the other allows the derivation of the above results. The above theorems present a technique for computing the translational parameters up to a scale. First, one computes $M_\Omega(A_5)$ and $M_\Omega(A_6)$. If the lines joining $M_\Omega(A_5)$ and B_5 and the line joining $M_\Omega(A_6)$ and B_6 are parallel but not coincident, then the direction of the line is the translation. Otherwise, the translation T is the intersection of these two lines provided that they do not coincide.

We now consider the case in which these two lines coincide on the image plane. We will show that T could not be of the form $(a,b,0)^t$. Consider Figure 3 which depicts the motion of two rays where P_5^1, P_6^1 are backprojections of A_5, A_6 at time t_1 and Q_5^1 and Q_6^1 are two virtual points on Ω. After the rotation, these points are moved to points with superscipt 2 as depicted in the middle of Figure 3. If T is $(a,b,0)^t$, then the plane formed by O, P_5^2, P_6^2 is shifted to the plane formed by M, P_5^3, P_6^3. In particular, O is shifted to M. Now the projections of the two lines $P_5^3Q_5^3$ and $P_6^3Q_6^3$ coincide on the image plane. This implies that the camera origin O is on the plane defined by M, P_5^3, P_6^3. Clearly, this is impossible unless the translation will move O to any point on the line joining D and E. From above reasoning, we can then conclude that T lies on the projection line. Although T can not be determined yet, an additional constraint arises as a result of this situation. From Figure 3, the orientations of the plane defined by the two rays are known before and after the motion if the the line joining M_Ω (A_5) and B_5 coincides with the line joining M_Ω (A_6) and B_6. Thus the following theorem holds:

Theorem 4:

If the line joining M_Ω (A_5) and B_5 coincides with the line joining M_Ω (A_6) and B_6, then R maps $(A_5 \times A_6)$ to $B_5 \times M_\Omega(A_5)$ assuming the magnitudes of both vectors are properly normalized and T lies on the line joining M_Ω (A_5) and B_5.

4.4. Deriving the Rotational Parameters

In the previous section, we derived either (i) $T = (t_x, t_y, 0)^t$ or (ii) $T = (t_1, t_y, 1)^t$. For the time being, the case described in Theorem 4 will be ignored and will become evident at the end of this section. In either case, we arbitrarily choose two independent vectors which are perpendicular to T. For case (i) one could choose $T_1^* = (-t_y, t_x, 1)^t$ and $T_2^* = (-\frac{t_y}{2}, \frac{t_x}{2}, 1)$. For case (ii), recall that there are two lines passing through $(t_x, t_y, 1)^t$, the following line equations on the image plane hold:

$$a_1 t_x + b_1 t_y + c_1 = 0 \qquad (11)$$

$$a_2 t_x + b_2 t_y + c_2 = 0 \qquad (12)$$

One can, therefore, choose

$$T_1^* = \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} \text{ and } T_2^* = \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix} .$$

We now construct two vectors T_1, T_2 such that $RT_1 = T_1^*$ and $RT_2 = T_2^*$. Since the procedure for constructing both remains the same, we shall temporarily omit T_2.

Choose two vectors p' and q' with third component 1, $p' \perp T_1^*$, and $q' \perp T_1^*$ where $\perp$ denotes perpendicularity. This can always be done because p' and q' are simply two points on the line defined by Equation (11). The purpose of choosing the third component to be 1 is to place p' and q' on the image plane. Now let $M_\Omega^{-1}(p') \equiv p$ and $M_\Omega^{-1}(q') \equiv q$, one has the following equations:

$$R\ z_p\ p = z'_p\ p' - T \qquad (13)$$
$$R\ z_q\ q = z'_q\ q' - T \qquad (14)$$

where $z_p\ p$ and $z_q\ q$ lie on Ω at time t_1 and $z'_p\ p'$ and $z'_q\ q'$ lie on Ω at time t_2. Notice that $M_\Omega^{-1}(.)$ reverses the roles of the two views.

Taking the scalar products of both sides of (13) and (14) with T_1^* gives

$$T_1^* \cdot R\ z_p\ p = 0 \quad \text{and} \quad T_1^* \cdot R\ z_q\ q = 0$$

Thus $T_1^* \perp Rp$ and $T_1^* \perp Rq$. It is also clear that

$$R(p \times q) \perp Rp \text{ and } R(p \times q) \perp Rq \quad ;$$

therefore,

$$T_1^* = R(p \times q) \text{ or } T_1^* = -R\ (p \times q)$$

assuming that T_1^* is normalized to have the same length as $p \times q$.

Whether $T_1^* = R\ (p \times q)$ or $T_1^* = -R(p \times q)$, one notices that

$$\begin{aligned} T_1^* \cdot R\ z_1 A_1 &= \pm R(p \times q) \cdot R\ z_1 A_1 \\ &= \pm z_1 (p \times q) \cdot A_1 \\ &= z'_1 B_1 \cdot T_1^* \quad . \end{aligned}$$

Thus

$$\frac{z'_1}{z_1} = \frac{\pm (p \times q) \cdot A_1}{B_1 \cdot T_1^*} \quad .$$

Since $\frac{z'_1}{z_1}$ must be positive, the choice of sign can be resolved.

Thus $T_1^* = R\ (p_1 \times q_1)$, and $T_2^* = R\ (p_2 \times q_2)$. Now we have

$$\begin{aligned} R:\ & p_1 \times q_1 \rightarrow T_1^* \\ & p_2 \times q_2 \rightarrow T_2^* \\ & (p_1 \times q_1) \times (p_2 \times q_2) \rightarrow T_1^* \times T_2^* \cdot, \end{aligned}$$

so R is completely determined. For the case described in Theorem 4, we have $T_1^* = R\ (p_1 \times q_1)$ but do not have $T_2^* = R\ (p_2 \times q_2)$. However an additional constraint that maps $(A_5 \times A_6)$ to $B_5 \times M_\Omega(A_5)$ exists, thus one could determine R also.

4.5. Deriving the Depth

Recall that

$$z'_i B_i = R\ z_i A_i + T$$

To find z_i, one chooses a vector B_i^* perpendicular to B_i and takes a scalar product with both sides of above equation. This leads to

$$z_i = \frac{-T \cdot B_i^*}{A_i \cdot RB_i^*}$$

To find z'_i, one first chooses a vector A_i^* perpendicular to A_i. Next one takes the scalar product of RA_i^* with both sides of above equation. This leads to

$$z'_i = \frac{R\ A_i^* \cdot T}{R\ A_i^* \cdot B_i}$$

5. SIMULATION RESULTS

Before running simulations to investigate the robustness of our algorithm, we briefly discuss the criterion. We will evaluate the relative mean error of the translational vector in terms of the first two components since the third one is normalized to 1. We will compute the estimate $\hat{R}$ of the rotation matrix R and evaluate $\| \hat{R} - R \|$ in l_2 norm. Both the mean error and the standard derivation will be computed. Note that the following holds:

$$\frac{\| \hat{R}\ x - R\ x \|}{\| x \|} \leq \| \hat{R} - R \|$$

The geometric meaning is that the angle between $R\ x$ and $\hat{R}\ x$ (for every x) cannot exceed $2\ \sin^{-1} (\frac{\| R - \hat{R} \|}{2})$. This fact can be seen in Figure 4 where $\| R - \hat{R} \|$ is the length defined by $R\ x$ and $\hat{R}\ x$. In Table I, we use the angle between $R\ x$ and $\hat{R}\ x$ instead of the numerical norm. The technique used to estimate the norm $\|R - \hat{R}\|$ is based on Gerschgorin's circle theorem ([8],p.304) for eigenvalue and the theorem that "The norm of A is the square root of the largest eigenvalue of $A^T A$ " ([8], p.288).

All experiments assume the field of view of the camera is 60° and the object size subtends about 20° to 30°. The image is 512×512 pixels.

Eight simulations, each consisting of 48 experiments, were performed. The translation vector is (2, 2, 8) which corresponds to (0.25, 0.25, 1) when normalized; the rotation matrix is

$$R = \begin{bmatrix} 0.98525 & -0.17093 & 0.00779 \\ 0.17108 & 0.98482 & -0.02924 \\ -0.00267 & 0.03014 & 0.99954 \end{bmatrix}$$

which corresponds to 10 degrees of tilt, 10 degrees of slant and 10 degrees of rotation angle about the rotational axis. The results are listed in the Table 1. The first column of Table. lists the range of noise. For instance, 0-3 means that noise ranging from 0 to 3 pixels is uniformly added to the horizontal and vertical positions of the data independently. Thus the maximum error between the accurate data and the noisy data is $3\times\sqrt{2}\approx 4$ pixels. The largest noise in the last row is about 12 pixels. In the second column, we list the relative error in terms of translation vector. In the third column, the first entry is the mean error, the second is the standard deviation. For instance, the result of the first simulation says: If one applies the true R and the derived $\hat{R}$ to any vector x, then the angle between the resulting two vectors will not exceed 0.5 degrees on average with standard deviation 0.1 degrees.

For specificitiy, we list in Table 2 the error in $A_i's$ and $B_i's$ of a particular experiment chosen randomly from the simulations represented in the last row of Table 1. The translation derived is (0.284, 0.304, 1.0) as opposed to the correct (0.25, 0.25 ,1.0). The angle between estimated $\hat{R}x$ and Rx (for every x) is less than 4°. In terms of tilt, slant, and rotational angle, the estimated solution is (8°, 18°, 8°). The correct solution is

(10°, 10°, 10°).

6. Discussion and Conclusion

From our theory, two pieces of information are required for the unique recovery of structure from motion. One is the image motion of a plane and the other one is the image motion of points not on this plane. [9] shows that knowledge of the image motion of a plane would yield a finite number of solutions (two at most from simulation results). It is shown here that the second piece of information makes the recovery unique. By integrating these two pieces of information, not only we have presented a novel and robust technique but we have also shown a new way of examining the problem. To determine the image motion of any point on a plane, we have shown that four image correspondences are sufficient. For the information on points not on the plane, we show that two image correspondences suffice.

Several open questions remain: (i) Is "four image correspondences" a necessary condition for specifying the image movement of a plane? The answer is probably positive. However, to derive the constraints that arise by using three points would probably bring one to understand fully the feature based method. (ii) Could one reduce "two image correspondences" in the second piece of information to "one image correspondence"? The answer is probably negative. (iii) It is also desirable to formulate a least squares technique of this method by using more points when they are available to improve accuracy.

Longuett-Higgins enumerates the configuration of points in space that lead to the singularities of E in the 8-point algorithm. This means that one could not conclude the uniqueness of the recovery through this technique. For instance: (i) If any six of them lie on a conic, then the 8-point algorithm is defeated. From our result, one can show that there is a unique recovery for this case because there are six points lying on a plane and two points not on the plane. In fact, Tsai and Huang [9] also notes this corollary although it has to run the 8-point algorithm to derive the solution. (ii) If any seven of the the points lie in a plane then the 8-point algorithm is defeated. This conforms to our observation because one needs two other points not on the plane.

From this framework, one could easily understand the previous results in [2,3,6,9]. It also shows what simple mathematics can do for this complex problem.

References

1. Aggarwal, J.K. and Y.F. Wang, "Analysis of a Sequence of Images Using Point and Line Correspondences", *IEEE Int. Conf. on Robotics and Automation*, 1987, pp. 1275-1280.

2. Longuest-Higgins, H.C., "A Computer Algorithm for Reconstructing a Scene from Two Projections", *Nature*, Vol. 293, Septebmer 1981, pp. 133-135.

3. Longuet-Higgins, H.C., "The Reconstruction of a Scene from Two Projections - Configurations That Defeat the 8-Point Algorithm", *Proceedings of the First Conference on ARtificial Intelligence Applications*, Denver, December 5-7, 1984, pp. 395-397.

4. Nagel, Hans-Hellmut and Bernd Neumann, "On 3D Reconstruction from Two Perspective Views", *Proc. Int. Joint Conf. Artificial Intelligence*, Vancouver, Britich Columbia, 1981, pp. 661-663.

5. Nagel, H.H., "Image Sequences - ten (octal) Years - from Phenomenology Towards a Theoretical Foundation", *International Conference on Pattern Recognition*, Paris, 1986, pp. 1110-1112.

6. Roach, J.W. and J.K. Aggarwal, "Computer Tracking of Objects Moving in Space", *IEEE Transactions PAMI*, Vol. PAMI-1, No. 2, April 1979, pp. 127-135.

7. Schunck, B.G., "Image Flow: Fundamentals and Future Research", *Conference on Computer Vision and Pettern Recognition*, 1985, pp. 560-571.

8. Strang, Gilbert, *Linear Algebra and Its Applications 1980*, Academic Press, New York.

9. Tsai, R.Y. and T.S. Huang, "Estimating 3-D Motion Parameters of a Rigid Planar Patch, I", *IEEE Transactions ASSP*, Vol. ASSP-29, No. 6, December 1981, pp. 1147-1152.

10. Tsai, R.Y. and T.S. Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surface", *IEEE Transactions PAMI*, Vol. PAMI-6, No. 1, January 1984, pp.13-26.

11. Ullman, Shimon, *Image Understanding 1984*, Ablex Publishing Corporation, Norwood, New Jersey 07648.

12. Waxman, A.M., "An Image Flow Paradigm", *Proceeding of the Second Workshop on Computer Vision: Representation and Control*, Minneapolis, MN, April 1984, pp. 49-57.

13. Lee, C.H. and T.S. Huang, "Finding Point Correspondences and Determining Motion of a Rigid Object from two Weak Perspective Views ", Submitted to *Computer Vision, Graphics, and Image Processing*.

ERROR	TRANSLATION	ANGLE
0-1	0.9%	(0.5 ° , 0.1 °)
0-2	0.8%	(0.9 ° , 0.4 °)
0-3	1.0%	(1.2 ° , 0.5 °)
0-4	3.0%	(1.5 ° , 0.8 °)
0-5	2.0%	(1.9 ° , 0.8 °)
0-6	7.0%	(2.6 ° , 2.1 °)
0-7	5.0%	(3.2 ° , 1.6 °)
0-8	11.0%	(5.1 ° , 4.3 °)

Table 1

A_1	A_2	A_3	A_4	A_5	A_6
6.7	9.7	7.5	4.5	6.3	9.4
B_1	B_2	B_3	B_4	B_5	B_6
6.3	3.6	6.0	5.4	1.8	8.2

Table 2

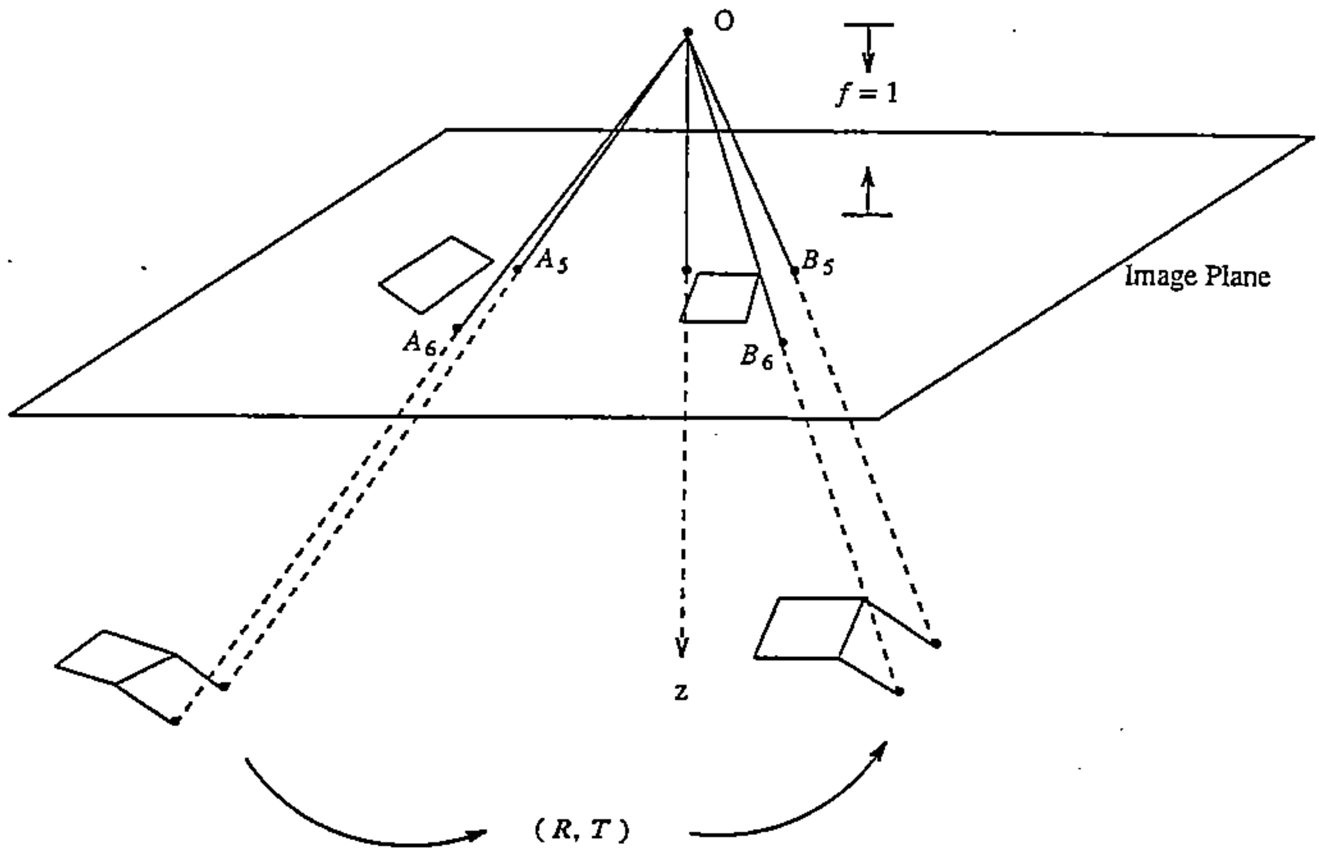

Figure 1. Imaging geometry and the problem.

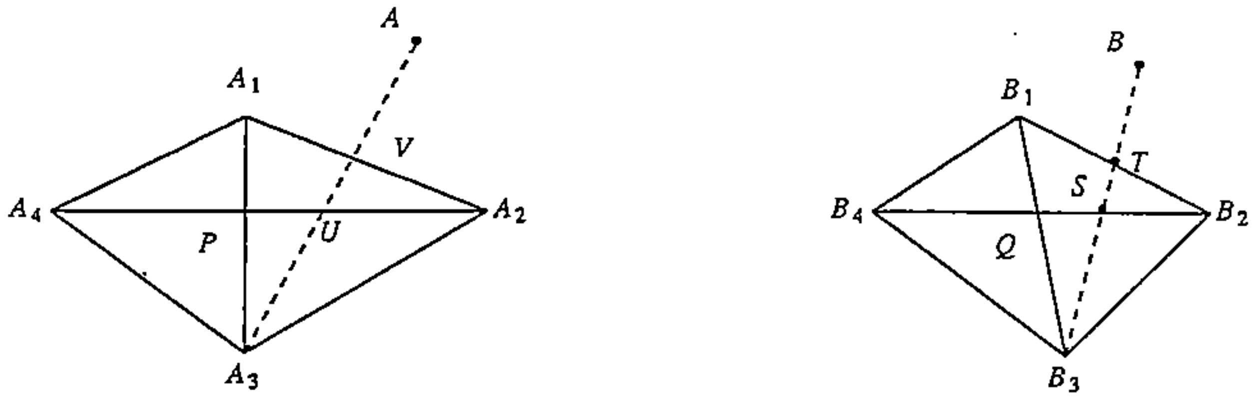

Figure 2: Compute $M_\Omega(A)$ by drawing.

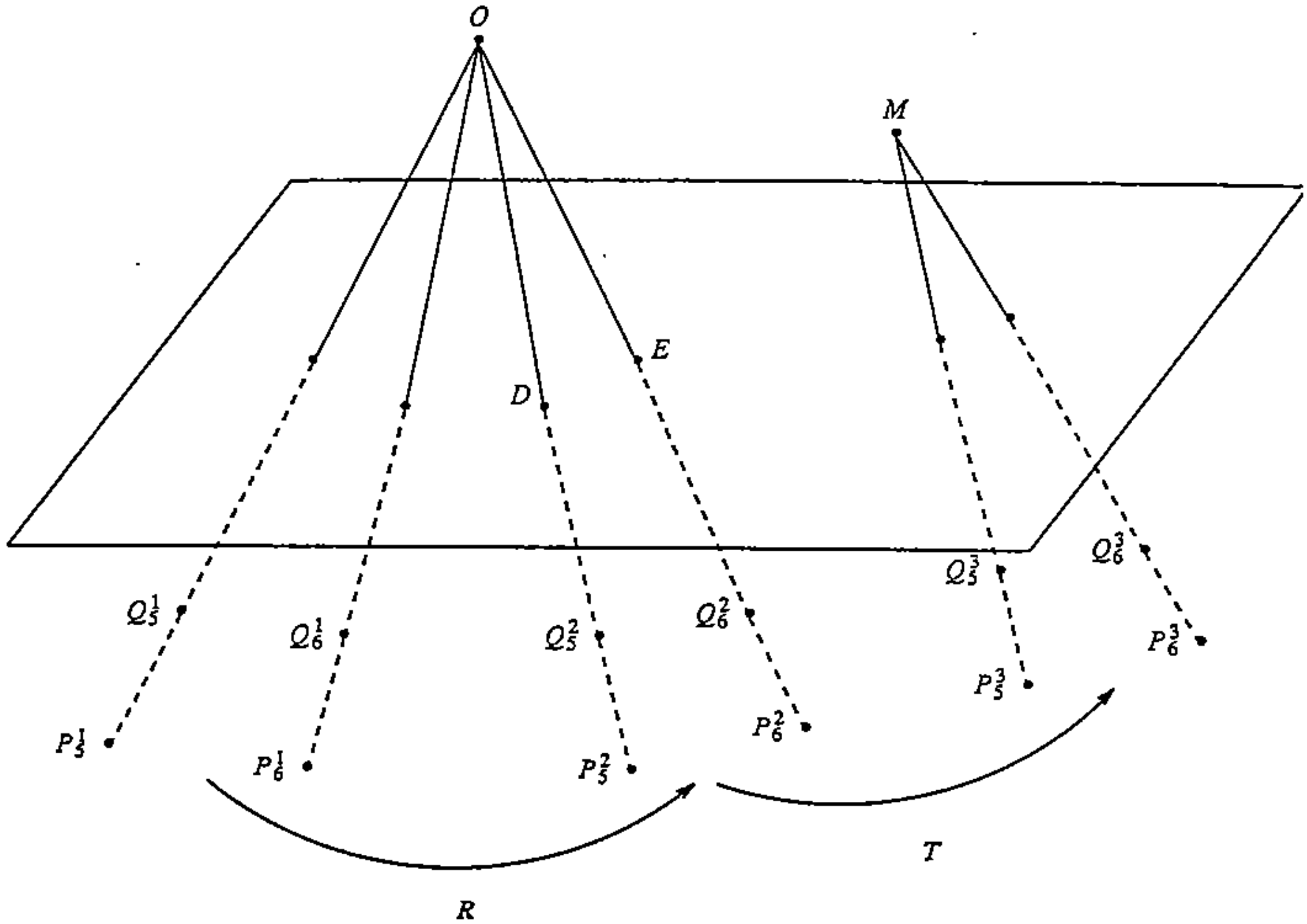

Figure 3.

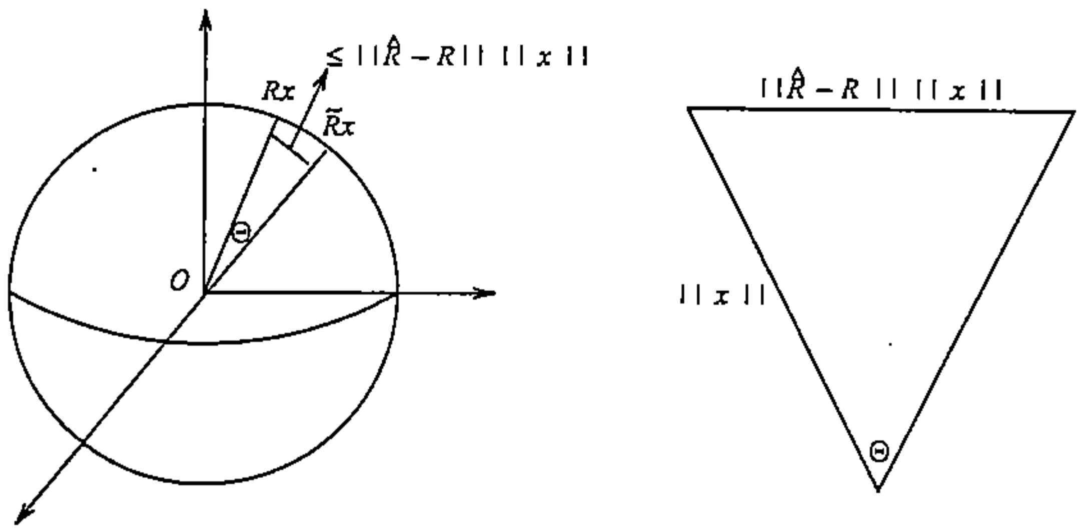

Figure 4