1985

Structure and Motion of a Rigid Object Having Unknown Constant Motion

Chia-Hoang Lee

Azriel Rosenfeld

Report Number:
85-548

STRUCTURE AND MOTION OF A RIGID OBJECT HAVING UNKNOWN CONSTANT MOTION

Chia-Hoang Lee
Azriel Rosenfeld

CSD-TR-548
December 1985

Structure and Motion of a Rigid Object Having Unknown Constant Motion

*Chia-Hoang Lee**

Department of Computer Sciences
Purdue University
West Lafayette, IN 47907

Azriel Rosenfeld

Center for Automation Research
University of Maryland
College Park, MD 20742

ABSTRACT

A method to estimate the underlying motion and the structure of an unknown rigid object is presented. This approach differs from previous techniques in the relationship assumed between successive motions. With this assumption, the number of feature points can be reduced to only two which would otherwise be not interpretable by previous method[4] even if finitely many frames could be used. A closly related assumption can be found in [1]. Several results are established: (1) There are two ambiguities when three frames are used. (2) There is a unique interpretation for the underlying motion and structure when a fourth frame is included. (3) The uniqueness theorem was proved by a geometric method which leads to an efficient, simple, and reliable algorithm. (4) Necessary criteria to reject the suitabiltiy of the assumption are stated for three and four frames. (5) Many test examples are reported with detailed implementation experience. They reveal that the algorithm is stable.

December 5, 1985

* The work was supported in part by the National Bureau of Standards under Grant 60NANB4D0053 while the author was at University of Maryland, and was completed at Purdue University. This report also appears as CAR-TR-157, Center for Automation Research, University of Maryland.

1. Introduction

The recovery of object structure from motion is an important task in either industrial applications or understanding of visual systems. One approach is to compute the optical flow of an imaged object and then estimate its structure [2,3,6,7]. Another approach which is more closely related to this paper uses feature points over several frames to find the 3D shape [1,4,5] based on rigidity constraints. One use of these approaches based on the rigidity assumption is to group (feature) points into different objects or different parts of an object and thus help segment the images.

One of the research directions in the past was to find a suitable combination of number of feature points and number of frames so that the observables are sufficient for determining the object structure (uniquely). Theoretically, finding such sufficient conditions is not a difficult task because relating the image coordinates to the object surface geometry is straightforward. The issue here is instead computational complexity and the reliability of the different proposed techniques. As to the uniqueness of the reconstructed structure, it is generally harder to address due to the nonlinearity of the equations.

Another interesting research issue is to develop techniques which can also have interpretations when only two feature points are available. Several attempts have been made along this line to account for human perceptual ability as noted in Johansson's experiments. In his experiments, only the major joints of a moving human body are visible which implies that two feature points are observable for each rigid part. A recent paper[1] gives a good exposition and many references.

2. Related Work

In an early computational study [4], it was found that four noncoplanar points in three views uniquely determine the relative positions of these four points. Although three views were used, the motion between the first two views and the motion between the next two views are irrelevant. As has been known for a long time, any rigid motion (between two frames) can be represented as a rotation by an angle θ about an axis through any point, followed by a translation. Thus the irrelevant successive motions may suggest the possibility that the rotation axis is changing, or the translation differs, or both.

Over a short time period, it is reasonable to assume that the direction of the axis remains fixed, as pointed out in [1]. Using this simple assumption, a different technique[1] can be found to make use of as few as two rigidly connected points while the "four points three views" approach is not applicable. Moreover, it is easy to show that the rigidity assumption in "four points three views" cannot interpret the structure of two feature points by simply including more frames. The key observation in [1] is that a feature point will trace out a circle in three dimensional space if the relative positions are used. Clearly, recovering the structure can be done from the projected circle which of course is an ellipse (we ignore the degenerate case). Since it takes five points to fit an ellipse, therefore this method needs an observation period of at least five frames.

Usually, these frames are taken at equally spaced time intervals. From this consideration, we will extend the fixed axis assumption to require that the rotation angle remain the same for the frame to frame motion and call it constant.

Based on this, we develop a simple new technique which establishes the relative structure of as few as two feature points and the motion parameters from only three views. Furthermore, given any three frames, a necessary condition to ensure that the object is undergoing a constant rigid motion is developed. In addition, uniqueness from four frames is also proved.

3. Problem Description

In this paper, we consider the following problem. A rigid object Ω_0 moving uniformly in three dimensional space is imaged at several equally spaced time instants. The task is to recover the underlying motion and the structure of Ω_0 from these frames which are the projections of Ω_0 . Usually it is impossible to achieve this task if no identifiable feature points are assumed. Thus we assume that there exist at least two feature points A_0 , B_0 of Ω_0 which are traceable (i.e., the correspondence of A_0, B_0 is established among these frames; see Fig. 1).

To make the problem more specific : A world coordinate system with unknown E as the origin is assumed. The rigid object Ω_0 undergoes a constant motion and is transformed to $\Omega_{i+1} = \mathrm{R}\Omega_i + \mathrm{T}$ in the next frame where R denotes the rotation; T denotes the translation; and Ω_i denotes the position of Ω_0 in the ith frame. In particular

$$A_{i+1} = \mathbf{R}A_i + \mathbf{T}$$
$$B_{i+1} = \mathbf{R}B_i + \mathbf{T}$$

where A_0 B_0 are the two feature points of the object Ω_0 in the three dimensional space. The issues to be explored here are (i) How can we recover the relative structure of A_0, B_0 if the object observed actually undergoes an unknown constant motion ? (ii) Does there exist a simple criterion to distinguish whether the object undergoes a constant motion or not ? (iii) How can we recover the underlying motion ? (iv) Are the recovered structure and motion unique ?

4. Structure and Motion Of The Feature Points

4.1. The Theory

From the previous section, we have

$$A_{i+1} = \mathbf{R}A_i + \mathbf{T} \tag{1}$$

$$B_{i+1} = \mathbf{R}B_i + \mathbf{T} \tag{2}$$

where **R** is the rotation; **T** is the translation; i denotes the ith frame.

Let

$\overline{A_i}$, $\overline{B_i}$ be the projections of A_i , B_i

$\omega_i = B_i - A_i = (a_i \;\; b_i \;\; c_i)$

$\overline{\omega_i} = \overline{B_i} - \overline{A_i} = (a_i \;\; b_i)$

$l \equiv |B_i - A_i|$ for all i ; l remains the same for all frames because of the rigid motion

$L = l^2$

$\alpha = \overline{\omega_2} \cdot \overline{\omega_1} - \overline{\omega_1} \cdot \overline{\omega_0} = a_1 a_2 + b_1 b_2 - a_0 a_1 + b_0 b_1$

$\gamma_i = |\overline{\omega_i}|^2 = a_i^2 + b_i^2$

Subtracting (1) from (2), we get

$$\omega_{i+1} = \mathbf{R}\omega_i \quad \text{for } 0 \leq i \leq 2$$

From $L = l^2 = \mid B_i - A_i \mid^2$, we get

$$c_i{}^2 = L - \gamma_i$$

and

$$c_i = \pm\sqrt{L - \gamma_i}$$

One observation here is that the angle between the successive vectors generated by a vector (with feature points as its two ends) which undergoes a constant motion is invariant. This fact, shown in the following lemma, leads us to an efficient and simple technique to recover the structure of the two feature points and the underlying motion.

Lemma 1: The inner product or the angle between the successive vectors ω_i is invariant if the rigid motion is constant during the observation period.

Proof:

$$\begin{aligned} \omega_{i+1} \cdot \omega_i &= \mathbf{R}\omega_i \cdot \mathbf{R}\omega_{i-1} \\ &= \omega_i \cdot \mathbf{R}^t\mathbf{R}\omega_{i-1} \\ &= \omega_i \cdot \omega_{i-1} \end{aligned}$$

where t denotes the transpose of a matrix. Furthermore, $\mid \omega_i \mid = l$ for all i which implies that the angle is invariant. Q.E.D.

Applying Lemma 1, we get the algebraic equation

$$a_1\, a_2 + b_1\, b_2 + c_1\, c_2 = a_0\, a_1 + b_0\, b_1 + c_0\, c_1 \tag{3}$$

and thus

$$\alpha = c_1(c_0 - c_1) \tag{4}$$

Raising both sides of (4) to the fourth power, we get

$$\alpha^4 - 2\alpha^2\,[(L - \gamma_1)(L - \gamma_0) + (L - \gamma_1)(L - \gamma_2)] + (L - \gamma_1)^2(\gamma_2 - \gamma_0)^2 = 0 \tag{5}$$

Rearranging (5) in terms of L , we get

$$u\ L^2 + v\ L + w = 0 \tag{6}$$

where

$$u = (\gamma_2 - \gamma_0)^2 - 4\alpha^2$$

$$v = 2\alpha^2\,(\gamma_0 + \gamma_1 + \gamma_2) - 2\gamma_1\,(\gamma_2 - \gamma_0)^2$$

$$w = \gamma_1^2\,(\gamma_2 - \gamma_0)^2 - 2\alpha^2(\gamma_0\gamma_1 + \gamma_1\gamma_2) + \alpha^4$$

Here (6) is a quadratic equation which is easy to solve. Thus there exist at most two solutions for L (L should also be $\geq \max(\gamma_0,\gamma_1,\gamma_2)$). Notice that we have only used three frames so far. Due to the possibility of a positive or negative sign of coefficient c_i, we may think that there exist 16 possible combinations of the placement of the vector in three frames (if two Ls are possible). Subsequently, this might generate several motions and relative structures.

In fact, the following lemmas shows that there exist at most two combinations (up to reflection in depth) which will be generated in most situations when three frames are used.

Lemma 2: If $\alpha \neq 0$ then there exists one combination of vectors in the three frames (up to reflection with respect to depth) for each possible L. If $\alpha = 0$ then there exist two combinations of vectors in the three frames (up to reflection with respect to depth) for each possible L. In other words, there exists a unique combination of the placement of a vector for each possible interpretation of L if the projection of $\overline{\omega_0}$ onto $\overline{\omega_1}$ is not the same as the projection of $\overline{\omega_2}$ onto $\overline{\omega_1}$.

Proof: The 8 combinations including reflection are the following:

(1) $(a_0,b_0,+\sqrt{L-\gamma_0})\ (a_1,b_1,+\sqrt{L-\gamma_1})\ (a_2,b_2,+\sqrt{L-\gamma_2})$

(2) $(a_0,b_0,+\sqrt{L-\gamma_0})\ (a_1,b_1,+\sqrt{L-\gamma_1})\ (a_2,b_2,-\sqrt{L-\gamma_2})$

(3) $(a_0,b_0,+\sqrt{L-\gamma_0})\ (a_1,b_1,-\sqrt{L-\gamma_1})\ (a_2,b_2,+\sqrt{L-\gamma_2})$

(4) $(a_0,b_0,+\sqrt{L-\gamma_0})\ (a_1,b_1,-\sqrt{L-\gamma_1})\ (a_2,b_2,-\sqrt{L-\gamma_2})$

(5) $(a_0,b_0,-\sqrt{L-\gamma_0})\ (a_1,b_1,+\sqrt{L-\gamma_1})\ (a_2,b_2,+\sqrt{L-\gamma_2})$

(6) $(a_0,b_0,-\sqrt{L-\gamma_0})\ (a_1,b_1,+\sqrt{L-\gamma_1})\ (a_2,b_2,-\sqrt{L-\gamma_2})$

(7) $(a_0,b_0,-\sqrt{L-\gamma_0})\ (a_1,b_1,-\sqrt{L-\gamma_1})\ (a_2,b_2,+\sqrt{L-\gamma_2})$

(8) $(a_0,b_0,-\sqrt{L-\gamma_0})\ (a_1,b_1,-\sqrt{L-\gamma_1})\ (a_2,b_2,-\sqrt{L-\gamma_2})$

where (1)(8), (2)(7), (3)(6), (4)(5) are the reflections of each other.

They are again reflections of each other.

We next deal with the case of $\alpha = 0$.

From (3) or (4), we have

$$\alpha = c_1 (c_0 - c_2)$$

Since $\alpha = 0$, we get

(i) $c_1 = 0$

Here there are only four combinations left, which reduces to two combinations (up to reflection)

(ii) $c_0 = c_2$

Since the signs of c_0 and c_2 must be the same, again only four combinations are left which reduces to two combinations. Q.E.D.

The above lemma shows that we usually have a unique structure of feature points for each L and sometimes there are two ambiguities. In the following lemma the conditions which determine how many Ls can be obtained are stated.

Lemma 3: (i) If $u < 0$ then there must exist a unique L

(ii) If $u = 0$ then there exists at most one L

(iii) If $u > 0$ then there exist exactly two Ls or no L at all

where $u = (\gamma_2 - \gamma_0)^2 - 4\alpha^2$.

Proof: (i) Since the coefficient u of L^2 is negative, the parabola must be convex downward as depicted in Fig 2. Furthermore the fact that $f(\gamma_0) \geq 0; f(\gamma_1) \geq 0; f(\gamma_2) \geq 0$ can be verified easily from (4) and thus $\gamma_0,\gamma_1,\gamma_2$ must lie in the middle segment ef. Obviously, only one L will be greater than or equal to the maximum of $\gamma_0,\gamma_1,\gamma_2$.

(ii) Since $u = 0$, f(L) becomes a linear equation, hence has at most one solution.

(iii) Since $u > 0$, f(L) is convex upward as depicted in Fig 3.

The fact that $f(\gamma_0) \geq 0; f(\gamma_1) \geq 0; f(\gamma_2) \geq 0$ implies that $\gamma_0,\gamma_1,\gamma_2$ lie outside the segment gh. If one of $\gamma_0,\gamma_1,\gamma_2$ lies to the right of h then no solution exists; otherwise there will be exactly two solutions. Q.E.D.

The next lemma gives us a necessary condition to have a solution when the case of (ii) or (iii) of Lemma 3 is encountered.

Lemma 4: If $u \geq 0$, then the necessary condition to have a solution is $2\gamma_1 > \gamma_0 + \gamma_2$.

Proof: Let L^* be the minimum for f(L) as depicted in Figure 4.

From (5), we obtain

$$L^* = \frac{\gamma_1(\gamma_2 - \gamma_0)^2 - \alpha^2(\gamma_0 + 2\gamma_1 + \gamma_2)}{(\gamma_2 - \gamma_0)^2 - 4\alpha^2}$$

In order to have a solution, $\gamma_1 \leq L^*$ must hold. By comparing γ_1 and L^*, we

get $2\gamma_1 \geq \gamma_0 + \gamma_2$ as the necessary condition.

If $u = 0$ then the only root L which equals $-\frac{w}{v}$ must be $\geq \gamma_1$ which leads to $2\gamma_1 > \gamma_0 + \gamma_2$. Q.E.D.

The structure of the feature points may be ambiguous when three frames are used, as pointed out in Lemma 3. In the next Lemma, we show that the uniqueness of structure can be assured when another frame is included.

Lemma 5: The structure of the feature points can be uniquely determined if four frames are given.

Proof: As shown before, the translation component can be ignored by considering the vector between the two feature points. Since the underlying motion is assumed to be constant, the vector ω_0,ω_1,ω_2 will be transformed to $\omega_1 = \mathbf{R}\ \omega_0$, $\omega_2 = \mathbf{R}\ \omega_1$; $\omega_3 = \mathbf{R}\ \omega_2$ as depicted in Fig. 5 where the end point of vector ω_i is denoted by δ_i if we move one end point of each ω_i to the origin of our camera system. Since we can view **R** as a rotation by the angle θ about some axis with direction cosines $(n_1\ n_2\ n_3)$, the points δ_1, δ_2, δ_3, δ_4 must lie on a circle in space and the angle between $O\delta_i$ and $O\delta_{i+1}$ must be θ where O is the center of the circle and lies on the rotational axis.

The projections of the $\delta_i{}'s$ onto the image plane are denoted by $\overline{\omega_i}$. The task here (see Fig. 6) is to show that δ_0 can be uniquely determined (up to reflection)

from the observable $\overline{\omega_i}$'s. In the following, we give a geometrical proof:

Let

$O = (h, k, m)$ be the center of the circle

O_1 be the midpoint of segment $\delta_1\delta_2$

O_2 be the midpoint of segment $\delta_0\delta_3$

O_3 be the midpoint of segment $\delta_1\delta_3$

as depicted in Fig. 6b.

Let

$\overline{O} = (h, k)$ be the projection of O

P_1 be the midpoint of segment $\overline{\omega_1\omega_2}$

P_2 be the midpoint of segment $\overline{\omega_0\omega_3}$

P_3 be the midpoint of segment $\overline{\omega_1\omega_3}$

as depicted in Fig. 6c where $\overline{O}$ is not observable.

Clearly P_i is the projection of O_i for $1 \leq i \leq 3$, because the midpoint property is preserved by parallel projection. Since O lies on the intersection of the two lines formed by O_1O_2 and δ_2O_3, therefore $\overline{O}$ must lie on the intersection of the two dashed lines formed by P_1P_2 and $\overline{\omega_2}P_3$ as depicted in Fig 6d. If we extend $\overline{\omega_2}\overline{O}$ (as indicated in the dotted line) to $\overline{\omega_4}$ such that the lengths $\overline{\omega_2\mathrm{O}}$ and $\overline{\mathrm{O}\omega_4}$ are equal, then $\overline{\omega_4}$ must lie on the same ellipse which is the projection of the cir-

cle. Moreover, a similar technique can be applied to $\overline{\omega_1}$, $\overline{\omega_3}$, $\overline{\omega_0}$ (although it is unnecessary) to find other new points on the ellipse. Obviously, five points can uniquely determine the orientation of the circle. Now let $(n_1\ n_2\ n_3)$ be the unit vector along the rotational axis which is the orientation of the circle; then there must exist a scalar t such that

$$t(n_1, n_2, n_3) + (x_0-h\ ,\ y_0-k\ ,\ z_0-m\ ) = (x_0, y_0, z_0)$$

where $(x_0, y_0, 0)$ is the projection of δ_0; (x_0, y_0, z_0) is δ_0 ; $(x_0-h\ ,\ y_0-k\ ,\ z_0-m\ )$ is the vector $O\delta_0$; and $(h\ ,\ k\ ,\ m\ )$ is the center of the circle.

Clearly $z_0 - m$ can be found by taking the inner product of (n_1, n_2, n_3) and $(x_0-h\ ,\ y_0-k\ ,\ z_0-m\ )$ to be zero and t can be found by comparing the first two components. Therefore, z_0 can be obtained from the projection.

The degenerate case would occur if both n_1 and n_2 are zero simultaneously which gives scalar t infinitely many possible values. In this case, any depth can be assigned to the feature point since the depth does not have any effect on the 3D motion. Q.E.D.

The motion parameters which consist of the rotational axis and rotational angle can be found easily. In fact, the orientation (n_1, n_2, n_3) of the circle in the previous lemma is the rotational axis. Furthermore, the tilt of the rotational axis is already found when the center of the ellipse is derived by the technique of Lemma 5 because the rotational axis passes through the camera origin and the center of the circle lies on the axis. To be precise, the center of the projected

circle and the camera origin determine the projection of the rotational axis. The slant of the rotational axis can be found from the lengths of the two axes of the ellipse. The cosine of the rotational angle can be found from the inner product of $O\delta_0$ and $O\delta_1$, taking the norms into account. We have already shown how to derive $O\delta_0$ in Lemma 5. The same technique can be used for $O\delta_1$.

4.2. An Implementation Algorithm

There are many ways to implement the above theory. Foe example, fitting an ellipse to five points can be done by solving a linear system of equations or by solving a linear equation [9] (using four points to obtain four straight lines, and the fifth point to determine the weight). Each of these methods may lead to undesirable results due to noise in the inputs. (To illustrate this situation, imagine the task of determining the intersection of two almost parallel lines. It would be better if we could bypass this step.) In the following, we shall outline an approach that results in an elegant, efficient and reliable implementation.

Although we mentioned finding the fifth point to get the ellipse in Lemma 5, we actually do not need this information in our construction process. Other information such as the axes and the center of the ellipse is more useful and reliable, based on our experience with several implementation algorithms. We outline the construction steps below.

(1) Move one endpoint of the observables to the origin of the camera. We thus obtain $\overline{\omega_0}$, $\overline{\omega_1}$, $\overline{\omega_2}$, $\overline{\omega_3}$.

(2) Use the midpoint technique as described in Lemma 5 to find the center (xcen ,

ycen) of the ellipse.

(3) It follows that the directions of the two axes of the ellipse are (xcen , ycen) and (ycen , –xcen). Therefore, the tilt of the rotation axis is arctan (ycen/xcen).

(4) Find the coordinates of any two points from $\overline{\omega_0}$, $\overline{\omega_1}$, $\overline{\omega_2}$, $\overline{\omega_3}$ with respect to the new coordinate system defined by the two axes of the ellipse, with the center of the ellipse as the origin of the system. (The idea here is to ensure that the equation for the ellipse is canonical.)

(5) Use the formula $\frac{X^2}{A^2} + \frac{Y^2}{B^2} = 1$ and the coordinates of (4) to find A and B.

(6) The slant of the last component of the rotational axis is arccos (B/A).

(7) (n_1, n_2, n_3) is found from the tilt and the slant from (3) and (6).

(8) Scalar t in Lemma 5 can be found from $t \cdot n_1 =$ xcen

(9) The last component $z_0 - m$ of $O\delta_0$ can be found from $(x_0-h\,,\ y_0-k\,,\ z_0-m) \cdot (n_1, n_2, n_3) = 0$

(10) z_0 can therefore be found as described in Lemma 5, and the relative structures of the two feature points can thus be obtained.

(11) Repeat (9) for $O\delta_1$. Find the inner product of $(x_0-h\,,\ y_0-k\,,\ z_0-m)$ and $(x_1-h\,,\ y_1-k\,,\ z_1-m)$ divided by the norms. The value obtained is the cosine of the rotational angle.

(12) To get the rotation matrix **R**, we can use the following matrix [8]:

$$\mathbf{R} = \begin{bmatrix} {n_1}^2 + (1- {n_1}^2)\cos\theta & n_1 n_2(1-\cos\theta) - n_3\sin\theta & n_1 n_3\,(1-\cos\theta) + n_2\sin\theta \\ n_1 n_2\,(1-\cos\theta) + n_3\sin\theta & {n_2}^2 + (1-{n_2}^2)\cos\theta & n_2 n_3\,(1-\cos\theta) - n_1\sin\theta \\ n_1 n_3\,(1-\cos\theta) - n_2\sin\theta & n_2 n_3(1-\cos\theta) + n_1\sin\theta & {n_3}^2 + (1-{n_3}^2) \end{bmatrix}$$

5. Experimental Results

Computer simulations for the estimation of the motion parameters and structure of the feature points were conducted on many sets of test data incorporating noise of 1% , 5%, and 10% of the signal.

In all the following examples, the first point used as a feature point is chosen randomly in three dimensional space with respect to the camera coordinate system. The other feature point is taken to be the reference point at the origin of the camera system, since it is the vector between the two points, not the absolute positions of the two points, that is significant in this context. Next, the point is rotated for several frames according to some arbitrarily chosen motion parameters which consist of the tilt and slant of a rotational axis, and a rotational angle about the axis. Then the observables are set to the first two components of these coordinates with noise added to them. The motion parameters and relative structure are computed using the method described in Sections 4.1 and 4.2. The results are then compared with the underlying motion and structure. In our implementation algorithm, one of the key steps is to find the center of the ellipse. This is important because the tilt of the rotational axis as well as the directions of the two axes of the ellipse are related to the coordinates of the center. Subsequently, all reconstructions are based on this canonical coordinate system for the ellipse. To demonstrate this key step, instead of deriving the center from noise-corrupted observables, we impose the same percentage of error in the center of the ellipse. The results are surprisingly reliable. Several examples were used to reveal that getting reasonable performance from the algorithm requires more than

four frames if the center is calculated from observables.

Example: Eight test cases are reported here. The tilt and the slant of the feature point were chosen to be 10 degrees and 80 degrees, and the radius to be 20 (using spherical coordinates). The tilt and the slant of the rotational axis were chosen to be 30 degrees and 40 degrees, and the rotational angle about the axis was 20 degrees. The first row gives the reference knowledge. The other rows show the results using different versions of algorithm as explained below.

Frames	Noise	**Motion Parameters**			**Feature Point**		
-	%	Tilt	Slant	Angle	Tilt	Slant	Radius
-	-	30	40	20	10	80	20
4	0	30	40	20	10	80	20
4	1	29	44	23	9	88	19.89
4	5	27	46	37	9	90	20.65
4	10	25	46	49	8	90	21.6
8	5	27	35	11	9	72	21.6
8	10	25	28	3	8	59	25.0
10	5	27	34	16	9	70	20
10	10	25	27	10	8	57	25.6

Many other examples were run under the conditions of no noise and four frames; the results all shows that the error due to the finite precision of computer arithmetic is less than 0.01%. The implementation up to this stage is a straightforward translation of the method described in Section 4.2. We have found that

other implementations using the fifth point to fit the ellipse by solving the linear system of equations sometimes produces an undesirable result due to the conditioning of the matrix for some inputs (e.g., very small rotational angle, small depth component of the feature point, etc.). The same situation was found using the approach in [9].

Noise ranging from 1% to 10% was introduced in the next set of tests, and we also corrupted the center of the ellipse together with the observables. In fact, we have found that a straightforward approach using the center derived from the corrupted observables in our algorithm suffers from similar difficulties to those encountered in the other implementations mentioned above. The correctness of the measurement of the center is again subject to the conditioning of the matrix. We see that the the algorithm with slight modifications yields very good estimates. In fact, some tricks need to be incorporated in the reconstruction algorithm even if the noise is directly imposed on the center of the ellipse, since the inherent difficulty encountered in fitting a canonical ellipse will certainly take place for some inputs. These situations often arise for a very small rotational angle or small depth component. However, often information can be exploited to overcome this difficulty, such as that the length of the long axis of the ellipse is greater than the distance from the center to any point on the ellipse. As another illustration, information about a point which is very near the endpoint of the long axis of the ellipse can also be used to estimate a bound on the length of the axis. By exploiting such information, we were able to to get very good performance over a very large range of input cases using only four frames.

We show that only one combination exists for $\alpha \neq 0$.

From (3) or (4) we have

$$\alpha = c_1(c_0 - c_2)$$

or

$$\alpha = \sqrt{L - \gamma_1}\ (\pm\sqrt{L - \gamma_0}\ \pm\ \sqrt{L - \gamma_2}\ ) \tag{7}$$

There are six cases to be considered: (i) $\alpha>0$, $\gamma_0 < \gamma_2$ (ii) $\alpha>0$, $\gamma_0 > \gamma_2$ (iii) $\alpha>0$, $\gamma_0 = \gamma_2$ (iv) $\alpha<0$, $\gamma_0 < \gamma_2$ (v) $\alpha<0$, $\gamma_0 > \gamma_2$ (vi) $\alpha<0$, $\gamma_0 = \gamma_2$. We here deal with cases (i) and (iii); the other cases are similar.

Case (i): $\alpha>0$,$\gamma_0 < \gamma_2$

Since $\sqrt{L - \gamma_0} > \sqrt{L - \gamma_2}$, it is easily seen from (7) that (8) and (9) can possibly hold:

$$\alpha = \sqrt{L - \gamma_1}\ (\ \sqrt{L - \gamma_0} + \sqrt{L - \gamma_2}\ ) \tag{8}$$

$$\alpha = \sqrt{L - \gamma_1}\ (\ \sqrt{L - \gamma_0} - \sqrt{L - \gamma_2}\ ) \tag{9}$$

If only one of them holds, then we are done. As an illustration, suppose (8) is true; then c_1,c_0 must have the same sign, and c_1,c_2 must have the same sign also. This implies that the following combination and its reflection are the only placements:

$$(a_0,b_0, +\sqrt{L - \gamma_0})\ \ (a_1,b_1, +\sqrt{L - \gamma_1})\ \ (a_2,b_2, +\sqrt{L - \gamma_2})$$

or

$$(a_0,b_0,-\sqrt{L-\gamma_0})\ (a_1,b_1,-\sqrt{L-\gamma_1})\ (a_2,b_2,-\sqrt{L-\gamma_2})$$

If both of them are true:

Subtracting (9) from (8) gives

$$\sqrt{L-\gamma_1}\,\sqrt{L-\gamma_2}=0$$

Adding (8) to (9) gives

$$\sqrt{L-\gamma_1}\,\sqrt{L-\gamma_0}=\alpha\neq 0$$

Thus $L=\gamma_2$ since L cannot be equal to γ_1. Furthermore, c_0 and c_1 must have the same sign to make $\alpha>0$. This implies that the following two combinations are possible:

$$(a_0,b_0,+\sqrt{L-\gamma_0})\,(a_1,b_1,+\sqrt{L-\gamma_1})\,(a_2,b_2,0)$$

$$(a_0,b_0,-\sqrt{L-\gamma_0})\,(a_1,b_1,-\sqrt{L-\gamma_1})\,(a_2,b_2,0)$$

They are again reflections of each other.

Case (iii): $\alpha>0$, $\gamma_0=\gamma_2$

Obviously, only (8) can hold which results in the following combinations :

$$(a_0,b_0,+\sqrt{L-\gamma_0})\ (a_1,b_1,+\sqrt{L-\gamma_1})\ (a_2,b_2,-\sqrt{L-\gamma_2})$$

$$(a_0,b_0,-\sqrt{L-\gamma_0})\ (a_1,b_1,-\sqrt{L-\gamma_1})\ (a_2,b_2,+\sqrt{L-\gamma_2})$$

The rest of the cases use the technique of linear least square estimation to derive the center of the ellipse and the lengths of the two axes from the observables in more than four frames, because it was found that the information mentioned in the previous paragraph will not work when the center is calculated from four corrupted frames. The errors propagated to the center from the observables are sometimes unpredictable and not tolerable. In these cases, the number of additional constraints on the center of the ellipse obtained by adding more frames is more than the number of the frames added which needs to be noted.

As these examples show, the proposed reconstruction algorithm is very reliable and yields good results. Of course, if some of the noise consists of outliers then a more sophisticated approach needs to be sought (e.g. filter out the outliers or use a feedback approach).

6. Concluding Remarks

A method of estimating the underlying motion and the structure of a unknown rigid object has been presented. This problem has been extensively studied in the analysis of time-varying imagery. The difference of this method from previous ones [4,5] is that the successive motion is assumed to be constant for a short time period. A similar method is described in [1] where a fixed rotational axis was assumed. Although ours is a stronger assumption, it allows a weaker condition on the number of observable points, which would otherwise be impossible to interpret using previous methods even if as many frames as desired are allowed to be taken. These approaches can be considered as not only representing a trade-off but also as complementary because neither approach can be extended to handle the other case.

Several results have been established. First, the fact that the angle remains invariant between successive frames was observed. Next, at most two ambiguities (up to reflection in depth) will occur if three frames are used. Furthermore, a simple criterion for a unique solution is developed if only three frames are available. Last but not least, the uniqueness of the underlying motion and the structure was proved geometrically, which led us to construct a simple, efficient, and reliable algorithm. One result we did not include in this paper is a very simple necessary criterion to have a constant motion for four points. In fact, the parallelism between the line formed by frame 1 and frame 4 and the line formed by frame 2 and frame 3 serves as the criterion.

Many test results were reported along with detailed implementation experience. The algorithm shows stability and yields good performance with up to 10% noise. Compared with the tolerance of noise in many other reported approaches, this technique seems robust. One experience we have learned is that fitting an ellipse to several points is better done in the canonical coordinate system if we can find the center and the axes.

References

[1] Webb, J. A. and J. K. Aggarwal (1982),"Structure from Motion of Rigid and Jointed Objects", *Artificial Intelligence* **19**, 107-130.

[2] Horn, B.K.P and B.G. Schunck (1981),"Determining Optical Flow", *Artificial Intelligence* **17**, 185-204.

[3] Longuet-Higgins, H. C. and K. Prazdny (1981),"The Interpretation of Moving Retinal Images", *Proceedings of the Royal Society of London* **B208**, 385-397.

[4] Ullman, S. (1979), "The Interpretation of Visual Motion", MIT Press, Cambridge, MA.

[5] Tsai, R.Y. and T. S. Huang (1984), "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces", *IEEE Trans. on Pattern Analysis and Machine Intelligence* **PAMI-6**, 13-27.

[6] Waxman, A. M. and S. Ullman (1983), "Surface Structure and 3-D Motion from Image Flow: A Kinematic Analysis", University of Maryland, Center for Automation Research, CAR-TR-24.

[7] Hoffman D.D. (1982), "Interpreting Time-Varying Images: The Planarity Assumption", *Proceedings of the Workshop on Computer Vision: Representation and Control,* Rindge, New Hampshire.

[8] Rogers D. F. and J.A. Adams (1976), "Mathematical Elements for Computer Graphics", McGraw-Hill, New York.

[9] Faux, I.O. and M.J.Pratt (1980),"Computational Geometry for Design and Manufacturing", Halsted Press, New York.

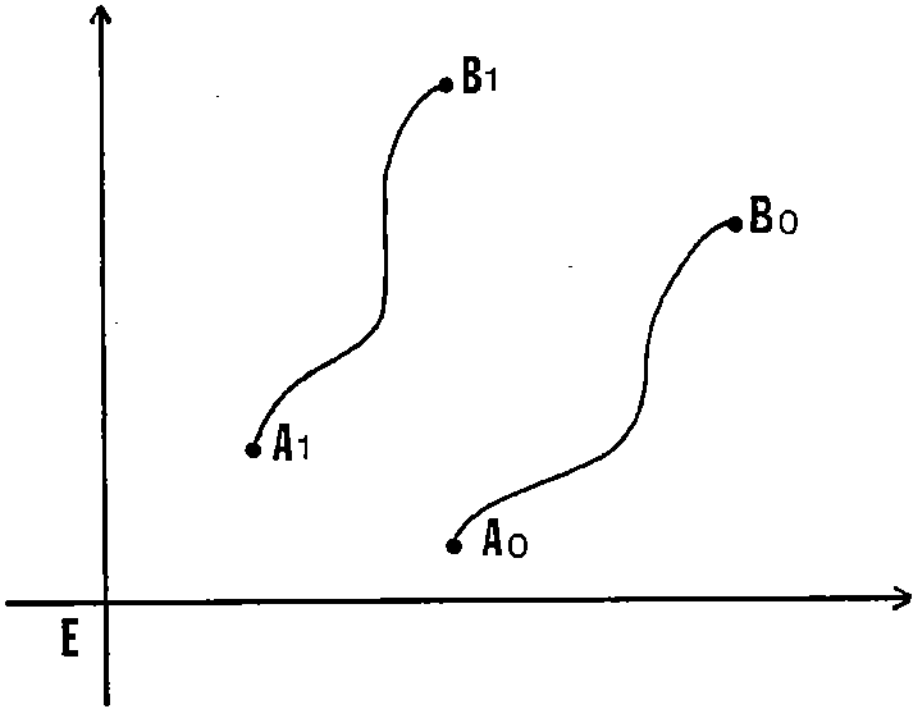

Figure 1.

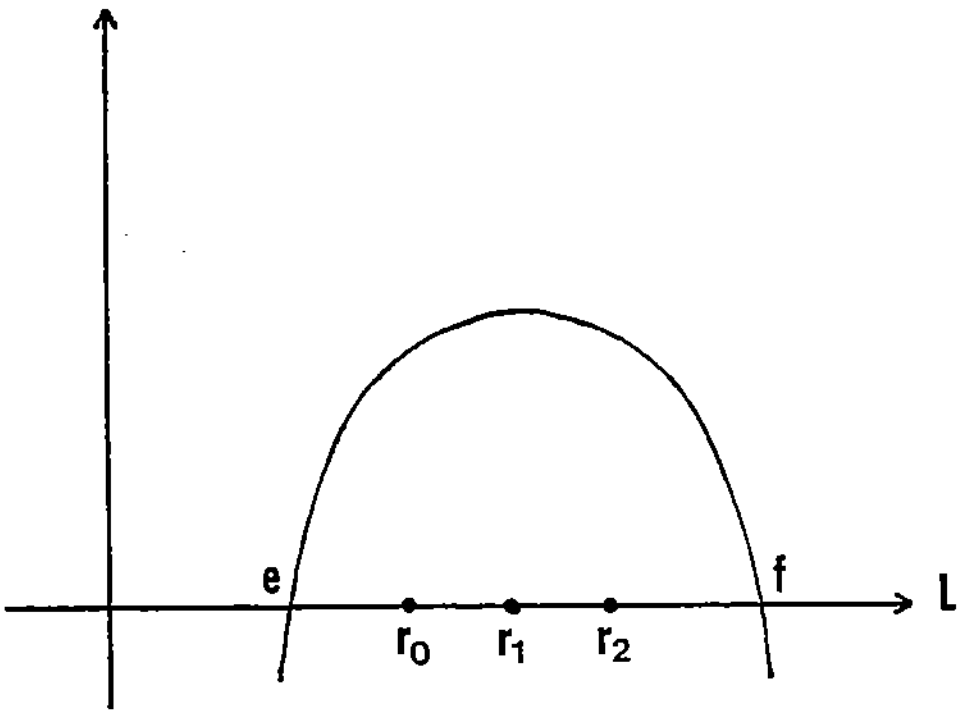

Figure 2: Parabola is convex downward; γ_i must lie in segment ef.

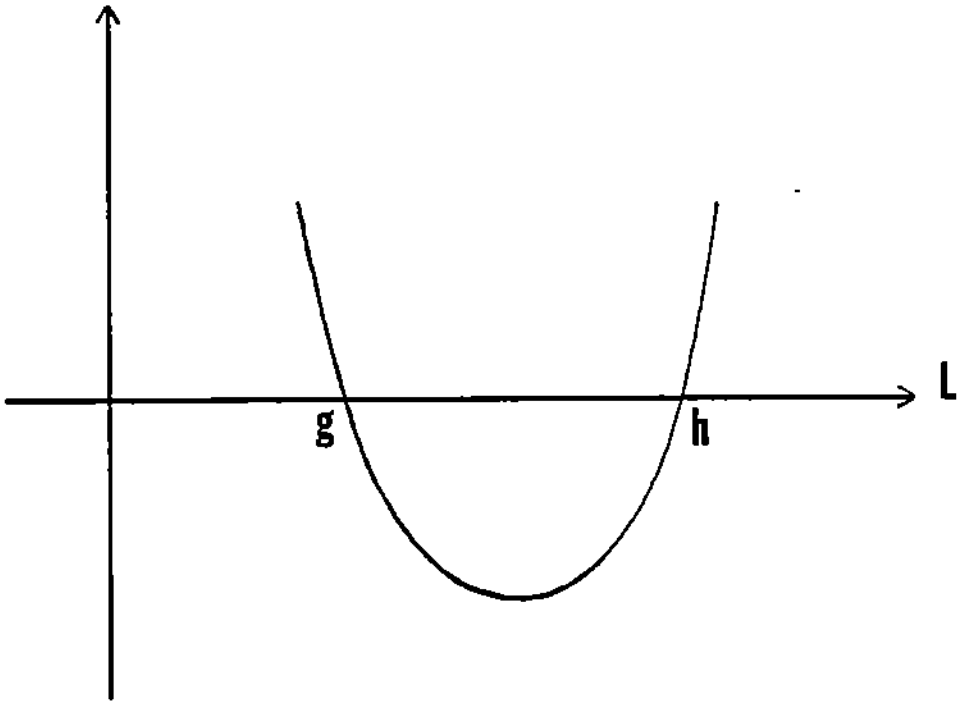

Figure 3: Parabola is convex upward; $\gamma_0\gamma_1\gamma_2$ cannot lie in the segment gh.

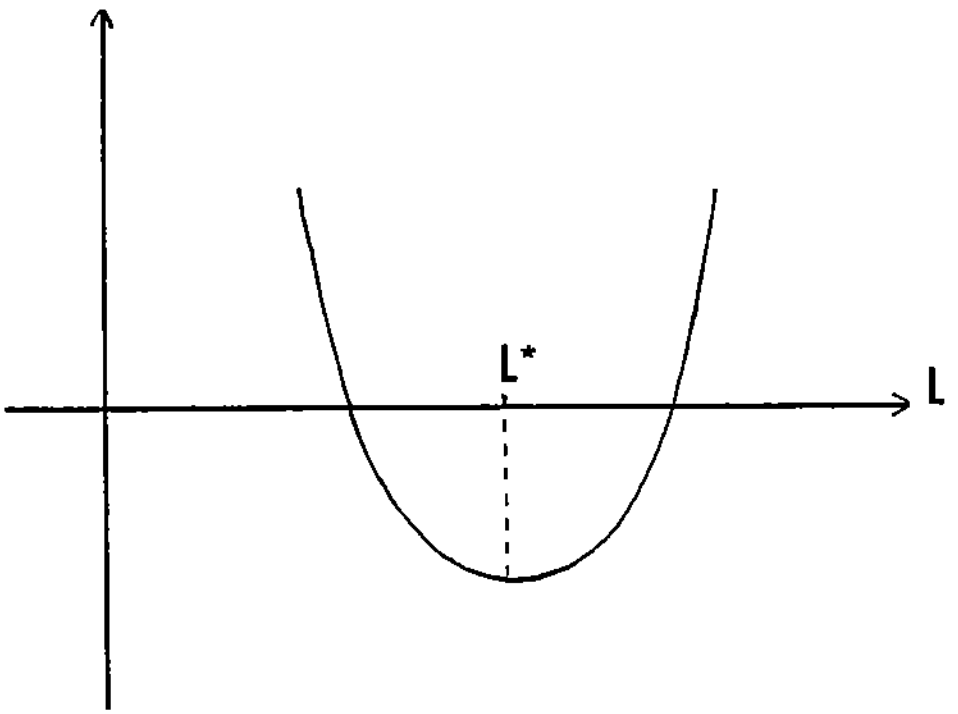

Figure 4: L^* assumes the minimum value.

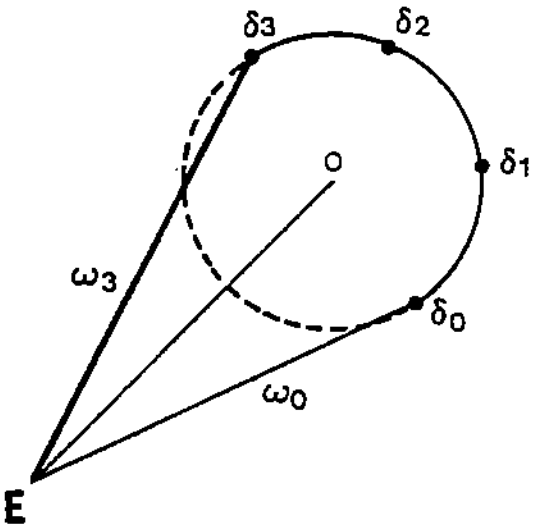

Figure 5: E is the origin of the camera system or reference point.

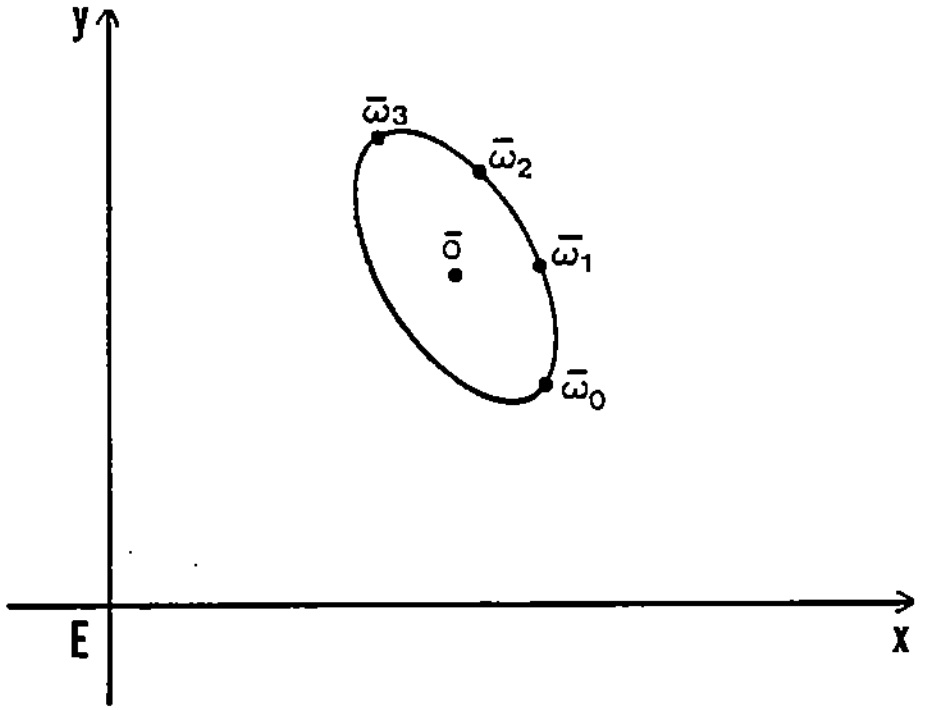

Figure 6(a): The projection of Figure 5.
The ellipse and $\bar{O}$ are not observable.
The task is to estimate σ_0 from $\overline{\omega_0\omega_1\omega_2\omega_3}$.

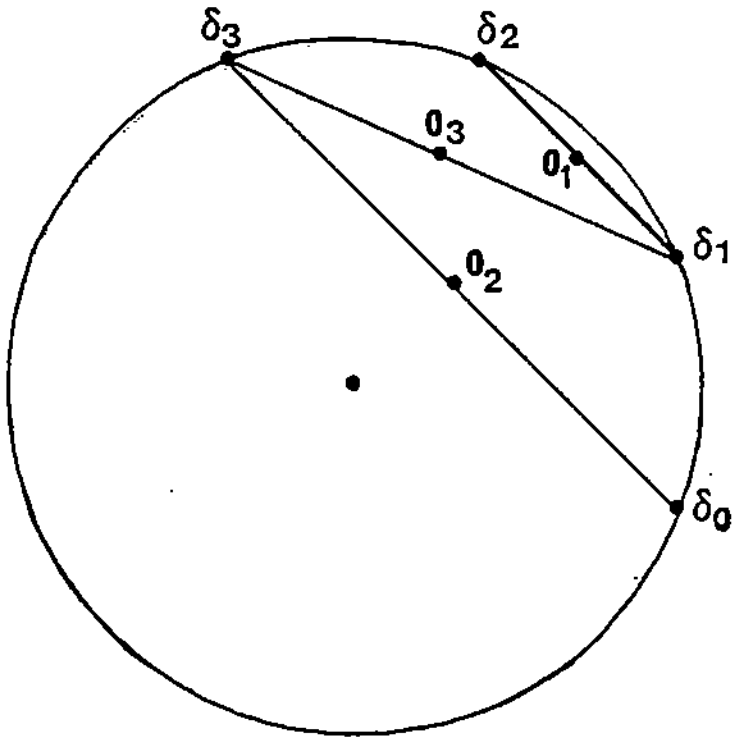

Figure 6(b): Imagine this figure in 3D space.

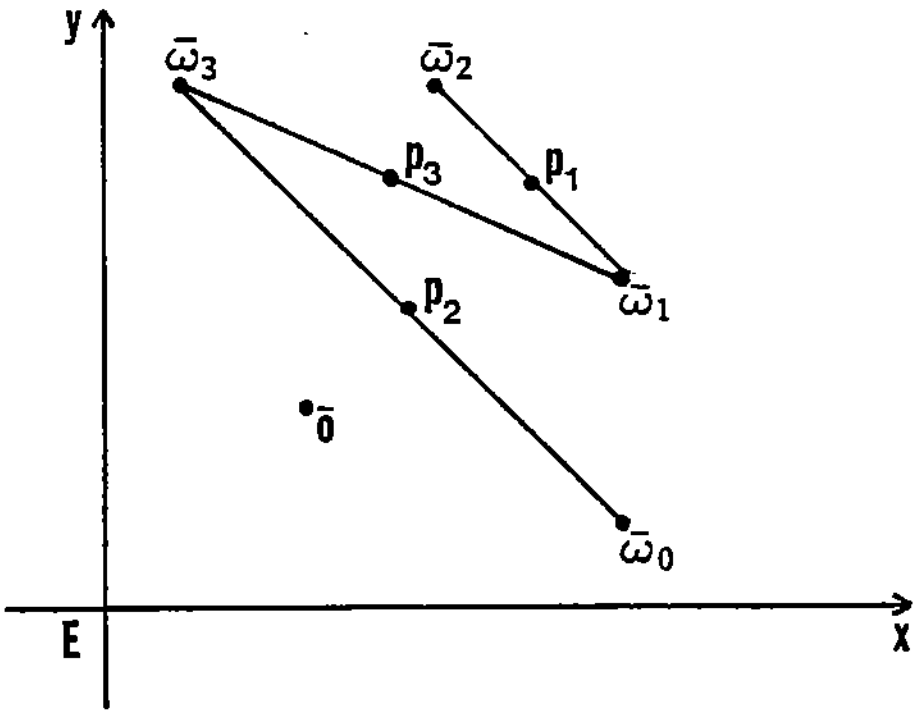

Figure 6(c): $\overline{O}$ is not observable.

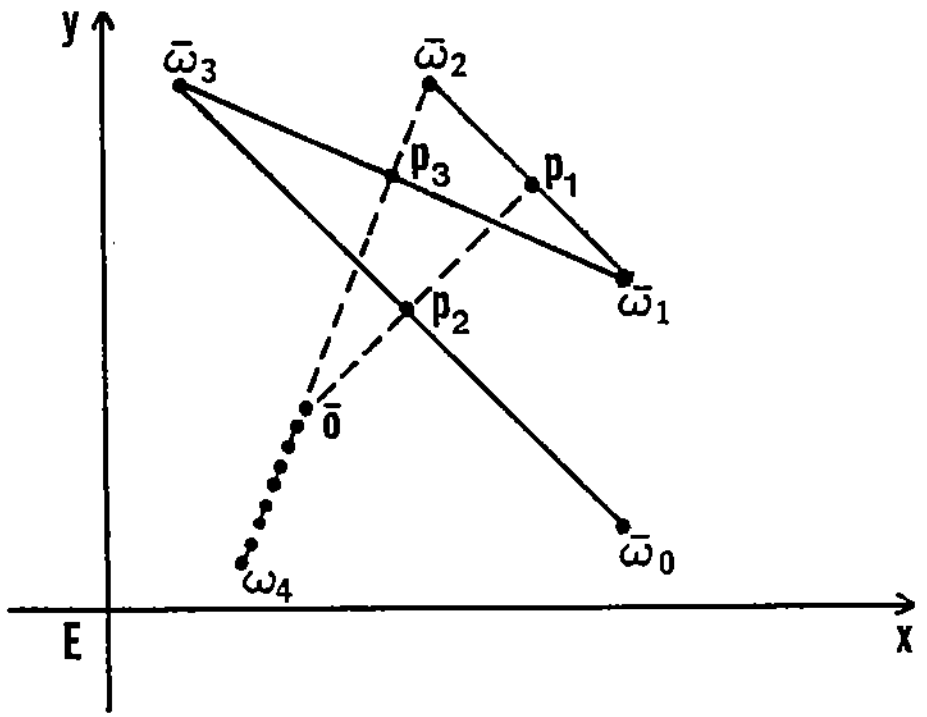

Figure 6(d): $\overline{O}$ is found and a new point $\overline{\omega_4}$ is created.