1986

On Correspondence, Motion, Scale and Structure of Two Views of a Scene

Chia-Hoang Lee

Report Number:
86-591

ON CORRESPONDENCE, MOTION, SCALE AND STRUCTURE OF TWO VIEWS OF A SCENE

Chia-Hoang Lee

CSD-TR-591
March 1986

ON CORRESPONDENCE, MOTION, SCALE and STRUCTURE OF TWO VIEWS OF A SCENE

Chia-Hoang Lee

Department of Computer Sciences

Purdue University

West Lafayette, IN 47907

ABSTRACT

Given two images of an n-points object which undergoes 3D rotation, translation and scaling. Our problems are (i) How can we match the corresponding elements in the two images due to the movement of the object? Can all the possible mapping be found ? (ii) What underlying motions and associated depth components of these points could account for the two images ? (iii) Can the object be recovered uniquely? This formulation of the problems referred to n-points problem is in the most general setting and does not assume attributes or features. A natural question to ask is whether an n-points problem is equivalent to a set of fewer-points problem. This paper presents a method which reduces an n-points problem to a set of 4-points problem. The effort of reduction takes $O(n)$ steps and it also takes $O(n)$ steps to construct all possible mappings of an n-points set from the solution to a 4-points problem. Other results include (1) Coplanarity condition of four points in two views. (2) Recovering the tilt direction of the rotational axis using four points in two views. (3) Recovering the scaling factor.

1. Introduction

One of the most fascinating abilities of human perception is to infer three dimensional (3D) structure from image data. Without any doubt, the ability is also a necessary feature for any flexible computer vision system. Although techniques of structured light, lasers, ultrasound are used successfully for 3D reconstruction in current machine vision systems, they are not flexible and have many restrictions. On the other hand, perception techniques based on shading, texture gradient, motion parallax, stereopsis and silhouettes are fast and flexible. It is thus interesting and important to study methods for obtaining 3D information from cues employed by human perception. In fact, this has become one of the major themes in vision research. For more details, interested readers are referred to [1] which contains a good collection of recent developments.

Among all the cues, motion parallax and stereopsis appear to contain the most information about the depth and have therefore attracted attention from most of researchers. However, techniques or algorithms developed using either motion parallax or stereopsis all face a basic "correspondence problem". This difficult problem is how to match the corresponding elements that occur in different images due to the movement of the objects or the observer.

1.1 An illustration of motion parallax and stereopsis

A general framework for approaches based on motion parallax is composed of three stages:

(1) Matching - a stage which determines the corresponding elements in different images or determines the optical flow when the motion is differential.

(2) Recovering the motion - a stage which recovers the motions of objects relative to the viewer based on the result of (1).

(3) Determining structure - a stage which determines the 3D structure of the scene.

In general, the solution for the third stage can be derived from the results of the first two stages very easily. Further, one usually treats the first two stages as separate research issues. Similary, approaches based on stereopsis would consist of the above steps except for the second one, which is assumed to be known in advance. In other words, the relative displacement of the two positions of the observer is known *a priori*. Therefore the essence of stereo algorithms lies in matching stage, with the step of deriving an object's depth being trivial.

As regards matching, conventional approaches [2] define a similarity measure based on a set of attributes and search for matches between windows from each frame or between feature points in the two frames. If the relative positions of the two views are known, as in stereo vision, then the search can be restricted to a one-dimensional space.

As regards motion recovery, one may rely on instantaneous velocity measurement, or on well-seperated feature points. Most studies derive a set of nonlinear equations which relate the image coordinates to the motion parameters and conclude that the number of feature points should be such that the number of equations is greater than or equal to the number of unknowns.

In this paper we present a theory that relates the correspondence, the motion and the structure, given two images which may have different scales. This theory is based on the assumption of parallel projection. We hope that there is a counterpart for perspective projection.

1.2 The Problem

Consider an object consisting of n points denoted by A_i $(0 \le i \le n-1)$ in 3D space. Let $\overline{A_i}$ be the projection of A_i into the image plane. We can rotate, translate and scale the object and observe the effect on the projections of the n points in the image plane. A precise mathematical model is as follows: Let $B_i = \sigma(R\ A_i + T) = \sigma\ R\ A_i + \sigma\ T$ where $0 \le i \le n-1$; R denotes 3D rotation; T denotes 3D translation; σ is an unknown nonzero constant. Define two sets, FIRST $\equiv \{ \bar{A}_i : 0 \le i \le n-1 ; \bar{A}_i$ is the first two components of $A_i \}$ and SECOND $\equiv \{ \bar{B}_i : 0 \le i \le n-1 ; \bar{B}_i$ is the first two components of $B_i \}$. Figure 1 depicts a situation in which FIRST consists of ten □'s and SECOND consists of ten O's. Our problems are (i) How can we match the corresponding elements in the two images due to the movement of the object? Can all the possible mappings be found ? (ii) What underlying motions (or discrete transformations) and associated depth components of these points could account for the two images ? (iii) Can the object be recovered uniquely ? I will refer to these problems on sets of n points as "n-point problems".

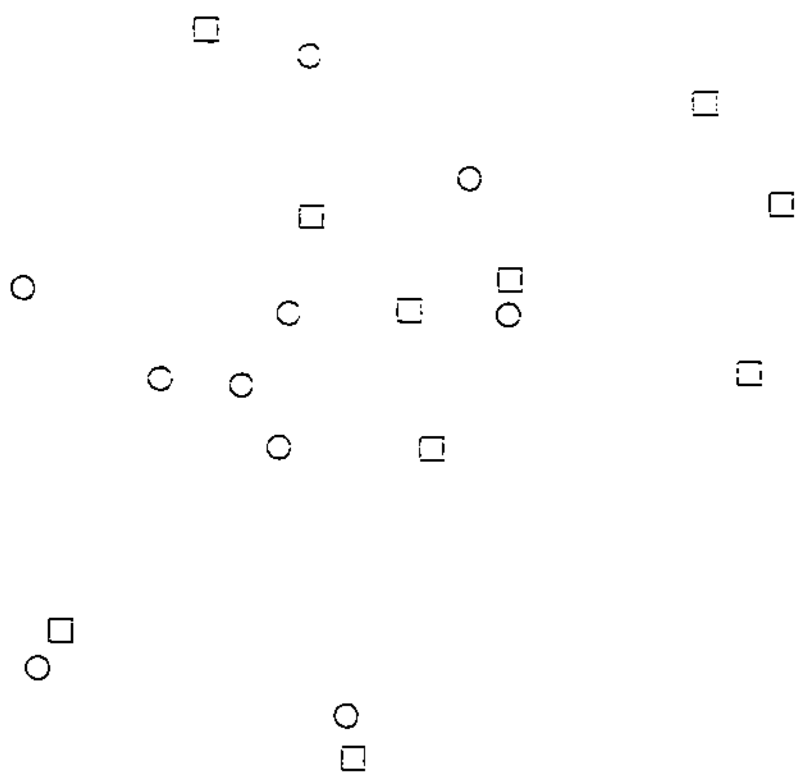

Figure 1

A more complicated problem is to consider several objects, instead of a single one, undergoing different motions. Thus FIRST and SECOND contain projections of several mixed objects. Another generalization of the problem adds uncertainty into the model.

For example, the positions of the projections may be corrupted by noise, or some points may show up in one image but not in the other due to occlusion or other reasons. Further, a curve segment or a region may replace a point in the problem.

As for problem (i), there are n! different one to one mappings between the two images. The computational complexity of examining each mapping becomes intolerable as n becomes large. Further. there seems to be no systematic method of rejecting an inconsistent mapping. In general, a unique solution for the possible mappings between FIRST and SECOND does not exist. Consider several points, uniformly separated around a circle in space. undergoing an arbitrary motion: then many mappings are possible. It is also apparent that a solution for problem (iii) not possible in general. As a trivial example, consider the case of two points; then there are infinitely many objects which are consistent with the images. It is, however, not clear what happens if the number of points is increased.

It is natural to ask if an n-points problem can be reduced to a set of fewer-points problem. There are two aspects of this question. The first aspect is the reduction step which enables one to study n-points problem on more manageable sets without affecting the answers. The second aspect is to analyze these problems on small sets. In this paper, we mainly discuss the reduction step along with several interesting observations about small-set problems. As an illustration, given the correspondence between two subsets of FIRST and SECOND shown in Figure 2, can all the consistent mappings between FIRST and SECOND be found subject to this constraint? What are the effects on the underlying motions and structures due to the reduction?

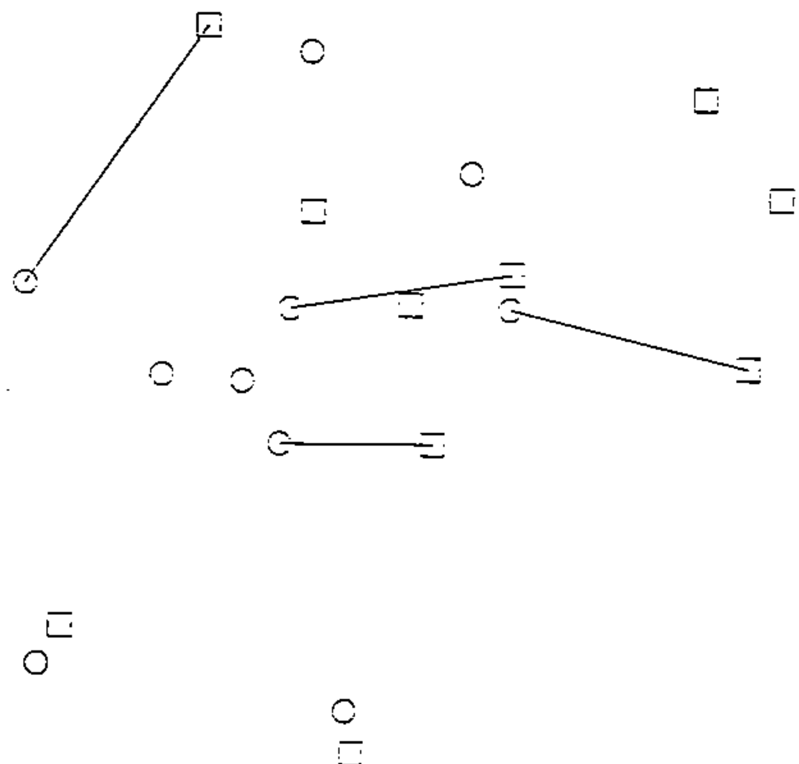

Figure 2

2. Theory

In the following subsections we derive several results for 4-point problems, assuming the correspondence is given. Next the relationship between n-point problems and a set of 4-point problems is analyzed based on these results.

Since the correspondence between the four points is given, one of the points will be chosen as the reference point A_0 and as the origin of the coordinate system. The coordinates of the other points will all be referred to A_0. By doing this, the rotational axis is adjusted to pass through A_0 without affecting the rotational matrix R. Further the translation T becomes zero because relative displacement instead of absolute position is used. The argument is as follows: Let FIRST and SECOND be given, and let A_0 be the reference point. Using A_0 as the origin of the coordinate system, we have, for $1 \le i \le n-1$,

$$\begin{aligned} B_i - B_0 &= \sigma (R \; A_i + T) - \sigma (R \; A_0 + T) \\ &= \sigma R \; (A_i - A_0) \end{aligned}$$

Renaming $B_i - B_0$ as B_i and $A_i - A_0$ as A_i, we have $B_i = \sigma R \; A_i$ for $1 \le i \le n-1$ with the understanding that A_0 is also an object point. Now we will assume that the correspondence between $\{\bar{A}_i : 1 \le i \le 3\}$ and $\{\bar{B}_i : 1 \le i \le 3\}$ is established and thus A_0 is a fixed point.

2.1 Notation

$A_0 \equiv O$, A_1, A_2, A_3 -- four points in the scene

$\bar{A}_0 \equiv O$, $\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$ -- projections of the A_i

O, B_1, B_2, B_3 -- same four points in the scene after motion and scaling

O, $\bar{B}_1$, $\bar{B}_2$, $\bar{B}_3$ -- projections of the B_i

Lower case Greek letters are used as scalars.

The rotational matrix denoted by R is written as

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = [r_1 \;\; r_2 \;\; r_3]$$

where r_i is the ith column of R. We will denote $(r_{13}\ r_{23})^t$ by l_1 and $(r_{31}\ r_{32})^t$ by l_2 where t denotes transpose.

It was shown in [5] that the rotational matrix has the following form:

$$R = \begin{bmatrix} n_1^2 + (1-n_1^2)\cos\theta & n_1n_2(1-\cos\theta) - n_3\sin\theta & n_1n_3\,(1-\cos\theta) + n_2\sin\theta \\ n_1n_2\,(1-\cos\theta) + n_3\sin\theta & n_2^2 + (1-n_2^2)\cos\theta & n_2n_3\,(1-\cos\theta) - n_1\sin\theta \\ n_1n_3\,(1-\cos\theta) - n_2\sin\theta & n_2n_3(1-\cos\theta) + n_1\sin\theta & n_3^2 + (1-n_3^2)\cos\theta \end{bmatrix}$$

where $(n_1\ n_2\ n_3)$ is the rotational axis ; θ is the rotational angle; and the tilt direction of the rotational axis is given by $(n_1\ n_2)$.

2.2 Degenerate Motions

We shall call a motion of the following form as degenerate:

$$R = \begin{bmatrix} \cdot & \cdot & 0 \\ \cdot & \cdot & 0 \\ 0 & 0 & \pm 1 \end{bmatrix}$$

This class of motions can arise in several ways:

(1) The optical axis is the rotational axis, i.e., every point rotates about the viewing axis.

(2) The direction of the rotational axis can be oriented arbitrarily, but the rotational angle is a multiple of 360 degrees, which is equivalent to no motion.

(3) The rotational axis lies on the image plane and rotational angle is 180 degrees.

We see, in these cases of degenerate motion, that the projection plane remains the same for these frames. Obviously, these frames can be generated from a single 2D image through rotation or reflection or both. The physical meaning of degenerate motion will be more clear if we treat the object as stationary with the relative motion in space attributed to the observer. In degenerate motion, the observer does not change his viewing direction. The observables are essentially equivalent to a single image with regard to the structure of the object, such motion is called degenerate motion. Theorem 1 proved in Section 2.3, is stated here in order to further discuss the detection of degenerate motion.

Theorem 1: Let O, $\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$ and O, $\bar{B}_1$, $\bar{B}_2$, $\bar{B}_3$ be two projections of four points undergoing nondegenerate motion. Let $\delta \bar{A}_1 + \gamma \bar{A}_2 = \bar{A}_3$ hold for some scalars δ, γ. Then a necessary and sufficient condition for O, A_1, A_2, A_3 to be coplanar in space is that $\delta \bar{B}_1 + \gamma \bar{B}_2 = \bar{B}_3$. Here we assume that O will be chosen appropriately so that $\bar{A}_3$ can be generated by $\bar{A}_1$, and $\bar{A}_2$ (e.g. if the projections of three points are collinear, then the other point will be chosen as O, as shown in Figure 3).

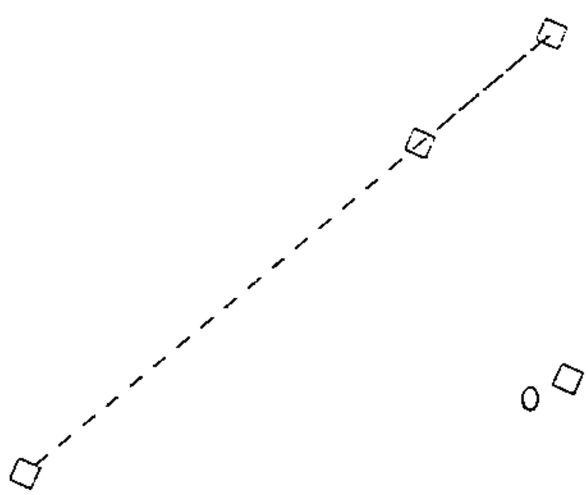

Figure 3

Criterion 1: Given a degenerate motion, then (i) $\|\bar{B}_i\| = \sigma \|\bar{A}_i\|$ for all i (ii) $\bar{B}_i \cdot \bar{B}_j = \sigma^2 \bar{A}_i \cdot \bar{A}_j$ for all i, j (iii) the condition of coplanarity holds.

Proof: Conditions (i) and (ii) are immediate since $\bar{B}_i = \sigma R^* \bar{A}_i$ and R^* is a 2 by 2 rotation. To show the condition of coplanarity, apply R^* to $\bar{A}_3 = \delta \bar{A}_1 + \gamma \bar{A}_2$. Q.E.D

Although these three conditions may not imply degenerate motion in all situations, we will use them as criteria for detection of degenerate motion. The only way to have these criteria satisfied for a nondegenerate motion is to distribute four coplanar points in a particular way, which is extremely unlikely. If the four given points are noncoplanar, then satisfying the criteria implies that the motion must be degenerate (see equation (4)). Most of our observations will relate to four noncoplanar points. Thus no false conclusions will be drawn here, because a nondegenerate motion of four noncoplanar points can never generate images satisfying the criteria.

2.3 Correspondence

In this subsection, we first introduce the concept of matching direction (similar to epipolar line), which is useful for finding possible matches in the second frame when a point in the first frame is given. This observation is an important step, though straightforward. Next, we derive a coplanarity condition and show that the matching direction can be derived if the given four points are not coplanar. Finally, we demonstrate an application of the matching direction leading to a reduction algorithm, described in Section 2.4, of an n-point problems to a set of 4-point problems.

Fact 1: Let R be a rotation depicted as above and (a, b) be an image point in the first frame; then the coordinates of the corresponding image point in second frame are $(a\ r_{11} + b\ r_{12} + s\ r_{13}, a\ r_{21} + b\ r_{22} + s\ r_{23})$ where s is a parameter. In other words, the corresponding point in the second frame can be any point of a line passing through $(a\ r_{11} + b\ r_{12}, a\ r_{21} + b\ r_{22})$ in direction (r_{13}, r_{23}).

Proof: Clearly, the coordinates of the corresponding point can be found to be $(a\ r_{11} + b\ r_{12} + s\ r_{13}, a\ r_{21} + b\ r_{22} + s\ r_{23})$ by applying R to (a, b, s) where s denotes depth. It is equivalent to say that the cooresponding point should lie on the straight line passing through $(a\ r_{11} + b\ r_{12}, a\ r_{21} + b\ r_{22})$ in direction (r_{13}, r_{23}). If we draw this line in the image plane. then any point on it could be a match. Q.E.D

From Fact 1, a family of parallel lines in the second frame is generated if we vary the given point in the first frame. On the other hand, Fact 1 can be applied to the second frame to generate a family of parallel lines in the first frame by *interchanging the roles of the first and second frames*. Further, the motion which transforms the second scene into the first scene is precisely the inverse or the transpose of R. Therefore, the parallel lines have (r_{31}, r_{32}) as their direction in the image plane; this is called the matching direction of the first scene with respect to the underlying motion. Now Theorem 1 stated above can be proved.

Proof of Theorem 1: The span, $\bar{A}_3 = \delta \bar{A}_1 + \gamma \bar{A}_2$, can always be assumed (see Figure 3). To establish sufficiency, let R^* be a principal minor of R as follows:

$$R^* = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$$

Since $B_i = \sigma R \; A_i$, assuming that s_i is the depth of A_i , we have

$$\bar{B}_i = \sigma R^* \bar{A}_i + \sigma s_i \, l_1 \qquad \text{for} \quad 1 \le i \le 3 \tag{1}$$

Write $\bar{A}_3 = \delta \bar{A}_1 + \gamma \bar{A}_2$. Applying σR^* to this relation and using (1), we obtain

$$\begin{aligned} \sigma R^* \bar{A}_3 &= \sigma \delta R^* \bar{A}_1 + \sigma \gamma R^* \bar{A}_2 \\ &= \delta (\bar{B}_1 - \sigma s_1 l_1) + \gamma (\bar{B}_2 - \sigma s_2 l_1) \\ &= \delta \bar{B}_1 + \gamma \bar{B}_2 - (\sigma \delta s_1 + \sigma \gamma s_2) l_1 \end{aligned} \tag{2}$$

Substituting (1) in the left hand side of (2), we have

$$\bar{B}_3 - \sigma s_3 \, l_1 = \delta \bar{B}_1 + \gamma \bar{B}_2 - (\sigma \delta s_1 + \sigma \gamma s_2) l_1$$

Hence

$$\sigma (\delta s_1 + \gamma s_2 - s_3) l_1 = \delta \bar{B}_1 + \gamma \bar{B}_2 - \bar{B}_3 \tag{3}$$

and

$$(\delta s_1 + \gamma s_2 - s_3) l_1 = 0 \tag{4}$$

Since l_1 is not a zero vector due to the assumption of nondegenerate motion, we have $(\delta s_1 + \gamma s_2 - s_3) = 0$. Hence, $A_3 = \delta A_1 + \gamma A_2$ whence coplanarity follows.

To show necessity, it is sufficient to observe that, since O, A_1, A_2, A_3 are coplanar, $\bar{A}_3 = \delta \bar{A}_1 + \gamma \bar{A}_2$ can be extended to $A_3 = \delta A_1 + \gamma A_2$. Now, applying rotation R and scale σ to the relation, we have $\delta B_1 + \gamma B_2 = B_3$. Consequently, we see that $\delta \bar{B}_1 + \gamma \bar{B}_2 = \bar{B}_3$. Q.E.D.

Theorem 1 provides a simple coplanarity criterion for four points from two image frames. Actually, the information in the fourth point becomes redundant after the coplanarity of the four points is recognized. Furthermore, the projection of any point lying on the plane containing these four points and its projection in the second frame after the motion can be easily established from the available observables.

Lemma 2 below demonstrates that the matching directions can be deduced if the four points are noncoplanar. This is a crucial step which enables the extension of the mapping between four points to n points as can be seen in Lemma 3.

Lemma 2: Given two different views of four points and a mapping between them, if they are noncoplanar then the matching directions in the two views are uniquely determined (up to sign) from the images.

Proof: As seen in Theorem 1, $(\delta s_1 + \gamma s_2 - s_3) l_1 = \delta \bar{B}_1 + \gamma \bar{B}_2 - \bar{B}_3$ holds. Since these four points are not coplanar, we have $\delta \bar{B}_0 + \gamma \bar{B}_1 - \bar{B}_2 \neq 0$. (Note: This condition is equivalent to nocoplanarity.) Thus we know l_1 up to an unknown constant. Furthermore, we can apply the same technique to derive l_2 up to an unknown constant by *interchanging the roles of the two frames*. Q.E.D.

Lemma 3: Given two different views of four points and a mapping between them, then the point corresponding to another object point in the first frame must lie on a line which can be derived in the second frame.

Proof: Let A be any non-feature point with s as its depth: let B be the point corresponding to A after the motion. Write $\overline{A} = \delta\overline{A}_1 + \gamma\overline{A}_2$. Applying R^* which is a minor of R and scale σ to the above relations, we have

$$\sigma R^* \overline{A} = \sigma \delta R^* \overline{A}_1 + \sigma \gamma R^* \overline{A}_2 \quad \text{and} \quad \overline{B} = \sigma R^* \overline{A} + \sigma s\, l_1$$

Hence

$$\overline{B} = \delta (\overline{B}_1 - \sigma s_1 l_1) + \gamma (\overline{B}_2 - \sigma s_2 l_1) + \sigma s\, l_1$$

Rearranging terms, we have

$$\overline{B} = \delta \overline{B}_1 + \gamma \overline{B}_2 + \sigma (s - \delta s_1 - \gamma s_2)\, l_1 \tag{6}$$

From this equation, we know that $\overline{B}$ must lie on a line passing through $\delta \overline{B}_0 + \gamma \overline{B}_1$ with direction l_1. Q.E.D.

2.4 Reduction Algorithm

In this subsection we discuss how the reduction is performed. We start by examining all possible mappings restricted to a three-point subset of FIRST and gradually extend these mappings over FIRST.

Choose a subset G3F of three noncollinear points from FIRST. Then there are $n(n-1)(n-2)$ mappings from G3F into SECOND. It is obvious that the n! mappings mentioned before are the straightforward extensions of these mappings to all of FIRST. Instead of examining each of n! mappings, our approach either extends each mapping from G3F into SECOND to a mapping from FIRST onto SECOND or rejects it. The first step is to extend the domain G3F of each mapping to four points, denoted by G4F, such that these four points are noncoplanar, and to spawn n-3 mappings (there is a situation in which e might need to spawn $O(n^2)$ mappings). The second step applies Lemma 3 to each of these n-3 mappings by considering points in FIRST - G4F. One may thus reject a mapping in one step or at worst in n-4 steps and have successful extensions to all of FIRST in n-4 steps. On the average, $O(n)$ steps are needed to establish consistent mappings and discard inconsistent ones. Further, we show later that $O(n)$ steps are needed to extend G3F to G4F using the algorithm below.

Definition: Let G3F = $\{O\ \overline{A}_1\ \overline{A}_2\}$ and let α be a mapping from G3F into SECOND. A planar transform induced by α, denoted by $planar_\alpha$, is a mapping from R^2 into R^2 and defined by

$$planar_\alpha\,(a\overline{A}_1 + b\overline{A}_2) = a\ \alpha(\overline{A}_1) + b\,\alpha(\overline{A}_2)$$

where a, b are real numbers.

The algorithm to extend G3F to four-element subset is given below; its correctness is shown afterwards.

Algorithm:

(*Initialization*):

Let P = { $P_1 \cdots P_{n-3}$} = FIRST - G3F;
Create n(n-1)(n-2) α_i's (mappings from G3F into SECOND) and let
$ALPHA = \{\alpha_i : 1 \le i \le n(n-1)(n-2)\}$

for $\alpha_i \in ALPHA$ do
    Q_{α_i} = SECOND - α_i(G3F);

    for $1 \le j \le n-3$ repeat
        $R_j := planar_{\alpha_i}(P_j)$
    until $R_j \in Q_{\alpha_i}$ is false;

    case:

    $j \ne n-3$: /* spawn n-3 mappings */
        Add P_j to G3F and spawn n-3 mappings
        by sending P_j to each element in
        SECOND - α(G3F) respectively;

    $j = n-3$: /* spawn $(n-3)(n-4)/2 + 1$ mappings */
        Record the "coplanar mapping" which sends P_i to R_i ;

        Add P_1 to G3F and spawn n-4 mappings by
        sending P_1 to each element in
        SECOND - α_i(G3F) $- planar_{\alpha_i}(P_1)$ respectively;

        Spawn n-5 mappings by sending P_1
        to $planar_{\alpha_i}(P_1)$ and P_2 to n-5 elements of
        SECOND - α_i(G3F) $- planar_{\alpha_i}(P_1) - planar_{\alpha_i}(P_2)$ respectively;

        Continuing in this manner, one spawns (n-4) + (n-5) + ...+1 mappings.

end

Further illustrations are given here. If there is at least one P_j such that $planar_{\alpha_i}(P_j) \in Q_{\alpha_i}$ is false then the first P_j is added to G3F. Because the extended mapping always interprets these four points as noncoplanar, Lemma 3 can be used. If $R_j \in Q_{\alpha_i}$ for $1 \le j \le n-3$, then we cannot spawn n-3 mappings by adding P_1 and by mapping it into any element in SECOND - α_i(G3F). The reason is that if P_1 is mapped into R_1 (one of the choices) then the extended mapping would interpret the four points as coplanar and Lemma 3 could not be used. Therefore only n-4 mappings are acceptable and would be spawned. As for the case in which P_1 is mapped to R_1 (which is of course one of the possible extended mappings, and we cannot simply discard it), we can consider possible correspondences of P_2. Clearly there are n-4 possible candidates, of which only one mapping (P_2 map into R_2) would interpret the five points (previous four points and P_2) as coplanar. Thus n-5 (instead of n-4) mappings would be spawned. Continuing in this manner for the other P_i's, we spawn altogether, in this case, $(n-3)(n-4)/2$ mappings.

There are $O(n^3)$ mappings to begin with and at worst $O(n^5)$ mappings would be generated if each mapping entered the second case (j = n-3) in the above algorithm. To generate a mapping for a four-point subset takes at worst $O(n)$ steps from a mapping for a three-point subset because we need only examine n-3 points in FIRST - G3F. The above-mentioned process assumes that all mappings for three points can be realized. Actually this is not the case; there are techniques [6] to examine whether a mapping is realizable or not. Thus one does not necessarily have to consider all $O(n^3)$ mappings. As for the extension over FIRST, after a mapping for four-points subsets G4F is constructed, one can apply Lemma 3 to the points in FIRST - G4F. The process is simple: Choose a point in FIRST - G4F; construct the corresponding line in the second frame; if no points in SECOND - α_i(G4F) lies on the line then reject α_i, otherwise record and continue. Obviously, it may take anywhere between 1 step and n-4 steps to reject a mapping, or up to n-4 steps to set up mappings from FIRST onto SECOND, where each step involes several additions/ subtractions and set membership operations.

Example: We demonstrate the extension of four points to n points. Consider the two images in Figure 2: FIRST = set of ten □'s and SECOND = set of ten O's. Choose 4 □'s and denote them by G4F. Clearly there are 10×9×8×7 = 5040 possible mappings.

We apply Lemma 3 to extend these mappings. A simple program can be written using four nested *for loops* with indexes from 1 to 10, but requiring that no two indexes should be the same. Surprisingly, it takes only one step to report failure for each of 5039 mappings and to report the correct and unique (in this example) correspondence.

2.5 Motion and Structure

This subsection reports four observations based on the assumption that a motion exists to account for a four-point mapping. The first two are related to the motion parameters. The third one recovers the scale and the last one addresses the effect of the reduction step.

Theorem 2: Given the mapping of four noncoplanar points in two frames, the direction of tilt of the rotational axis can only be $l_1 + l_2$.

Proof: Let R be the rotation matrix. We know from Theorem 1 that l_1, l_2 can each be derived up to an unknown constant. Let the unknown constants be c_1 and c_2. Because the norms of $c_1 l_1$ and $c_2 l_2$ must be the same to be consistent with the rotation matrix, we have

$$c_1^2 \, ||l_1||^2 = c_2^2 \, ||l_2||^2 \qquad \text{and} \qquad c_1^2 \, ||l_1||^2 \leq 1$$

Without loss of generality, we normalize l_1 and l_2 to be 1 and combine the two unknown constants c_1, c_2 into one unknown constant c which is a scalar less than 1. Therefore, we use $c \; l_1$ and $c \; l_2$ to represent the possible coefficients for (r_{13}, r_{23}) and (r_{31}, r_{32}).

Adding cl_1 and cl_2, we have

$$\begin{bmatrix} r_{13} \\ r_{23} \end{bmatrix} + \begin{bmatrix} r_{31} \\ r_{32} \end{bmatrix} = c \; (l_1 + l_2)$$

Thus

$$n_3\,(1-\cos\theta)\begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = c\begin{bmatrix} l_{11}+l_{21} \\ l_{12}+l_{22} \end{bmatrix}$$

Clearly $(l_{11}+l_{21},\, l_{12}+l_{22})'$, if not a zero vector, gives the tilt (i.e. the arctangent of the ratio n_2/n_1) of the rotational axis since $\cos\theta \neq 1$. Further the unknown signs of l_1 and l_2 may yield two possible tilts.

If $(l_{11}+l_{21},\, l_{12}+l_{22})' = (0\ ,\ 0)'$ then n_3 must be zero since $(n_1\ n_2)'$ cannot be a zero vector. due to nondegenerate motion. and $\cos\theta$ cannot be 1. We can thus derive $l_1 = c\,(n_2\ ,\ -n_1)$, and the tilt direction of the rotational axis is perpendicular to l_1. In this case. a unique solution is obtained. Q.E.D.

Theorem 3: The rotational angle (not 180 degrees) and the slant of the rotational axis must satisfy $n_3^2\tan^2\frac{\theta}{2} = \frac{\tan^2\tau - \Delta}{1-\Delta\tan^2\tau}$ where τ is the tilt of the rotational axis; Δ is the inverse of the product of the slopes of the matching directions. The limiting case (approaching infinity) is considered if divison by zero occurs.

Proof: Considering $\frac{r_{31}}{r_{32}}\cdot\frac{r_{13}}{r_{23}}$, we have

$$\frac{n_1^2\,n_3^2\,(1-\cos\theta)^2 - n_2^2\sin^2\theta}{n_2^2\,n_3^2\,(1-\cos\theta)^2 - n_1^2\sin^2\theta} = \frac{l_{21}}{l_{22}}\cdot\frac{l_{11}}{l_{12}} \tag{7}$$

Here we first assume that both r_{32} and r_{23} are nonzero and then discuss degenerate cases.

Let τ be the tilt of the rotational axis; then $\frac{n_2}{n_1} = \tan\tau$.

Thus equation (7) becomes

$$\frac{n_3^2\,(1-\cos\theta)^2 - \tan^2\tau\,\sin^2\theta}{\tan^2\tau\,n_3^2\,(1-\cos\theta)^2 - \sin^2\theta} = \frac{l_{21}}{l_{22}}\cdot\frac{l_{11}}{l_{12}} = \Delta$$

Expanding the above equation, we get

$$(1 - \Delta \tan^2\tau)\, n_3^{\,2}\, (1 - \cos\theta)^2 \;=\; (\tan^2\tau \div \Delta)\, \sin^2\theta$$

Rearranging the above terms, we have

$$n_3^{\,2} \tan^2\frac{\theta}{2} = \frac{\tan^2\tau - \Delta}{1 - \Delta \tan^2\tau} \tag{12}$$

Next we consider the two degenerate cases (a) (b) defined below:

(a) $r_{23} = 0, r_{32} \neq 0$: or $r_{23} \neq 0, r_{32} = 0$

Examining the coefficients of R, both situations yield

$$n_2^{\,2}\, n_3^{\,2}\, (1 - \cos\theta)^2 = n_1^{\,2} \sin^2\theta$$

and none of the terms in this equation are zero, which would otherwise violate the assumption stated in (a) that the other coefficient is not zero.

Rearranging this equation, we get

$$\tan^2\tau\, n_3^{\,2} \tan^2\frac{\theta}{2} = 1$$

Since $\tan^2\tau \neq 0$, we have

$$n_3^{\,2} \tan^2\frac{\theta}{2} = \frac{1}{\tan^2\tau}$$

This conclusion is exactly the same as relation (12) if Δ approaches infinity.

(b) $r_{23} = 0, r_{32} = 0$

This case yields $n_1 \sin\theta = 0$ and $n_2\, n_3\, (1 - \cos\theta) = 0$ which implies (i) $n_1 = 0$ and $n_2 n_3 = 0$ (ii) $\sin\theta = 0$ and $n_2 n_3 = 0$. (ii) is excluded because θ would be 180 degrees. Therefore we find that either the viewing direction (i.e. $n_3 = \pm 1$) is the rotational axis or the y-axis ($n_2 = \pm 1$) is the rotational axis. In the latter case, as in human binocular stereo geometry, formula (12) can still be used by letting both Δ and $\tan\tau$ approaching infinity. The former case, where the viewing direction is the rotational axis, is already excluded by the nondegeneracy assumption. Q.E.D.

Theorem 4: Given the mapping of four noncoplanar points in two frames, the scale σ can be recovered.

Proof: Let M be any point on the rotational axis and $\overline{M}$ be its projection. (We only need $\overline{M}$.) Further, let $\overline{M} = a\ \overline{A}_1 + b\ \overline{A}_2$. Clearly, the possible matches for $\overline{M}$ lie on the line denoted by m: $a\ \overline{B}_1 + b\ \overline{B}_2 + \sigma\,(t - as_1 - bs_2)\ l_1$ where t is the depth of M. If l_1 is not parallel to the tilt direction, then there is a unique intersection between m and the projection of the rotational axis. The difference between the intersection position and $\overline{M}$ is evidently due to σ and provides information for recovering it.

For the case of l_1 parallel to the tilt direction, which would yield overlap of m and the projection of the rotational axis, we note that this situation occurs only when the rotational angle θ equals 180 degrees. To see this, examine the coefficients of l_1 in (I). Since

$$l_1 = \alpha \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} + \sin\theta \begin{bmatrix} n_2 \\ -n_1 \end{bmatrix}$$

if l_1 is proportional to (n_1, n_2) one would have $\sin\theta = 0$, which is possible, or $n_2 = n_1$ and $-n_1 = n_2$, which leads to $n_1 = n_2 = 0$. The latter is impossible since the viewing axis cannot be the rotational axis. Now let $B_1 = C_1/\sigma$. If θ is 180 degrees, then the rotational axis intersects A_1C_1 at the midpoint. Therefore, the projection of the rotational axis intersects $\overline{A}_1\overline{C}_1$ at the midpoint which implies that the projections of $\overline{A}_1$ and $\overline{C}_1$ on the direction $(n_2, -n_1)$ are equal. Thus

$$\overline{A}_1 \cdot (n_2, -n_1) = \overline{C}_1 \cdot (n_2, -n_1) = \sigma\overline{B}_1 \cdot (n_2, -n_1) \quad \text{Q.E.D.}$$

The next theorem shows that if a motion can account for the mapping of four noncoplanar points and the mapping can be extended over FIRST onto SECOND, then the same motion can account for any extended mappings.

Theorem 5: Let Ω be a motion which can account for a mapping of four noncoplanar points. Then Ω can account for any extended mapping over FIRST onto SECOND.

Proof: Let $\bar{A} \in$ FIRST correspond to $\bar{B} \in$ SECOND under an extended mapping. From equation (6), we know that

$$\bar{B} = \delta \bar{B}_1 + \gamma \bar{B}_2 + \sigma (s - \delta s_1 - \gamma s_2) l_1$$

Since σ is recovered from the previous theorem and l_1 is known, we have $s - \delta s_0 - \gamma s_1$ as some constant. Thus s can be derived. Further (6) can be rewritten as $\bar{B} = \sigma R^* \bar{A} + \sigma s \; l_1$. This expression describes the transformation of A to B through R and σ. Therefore, if s is assigned as the depth of $\bar{A}$, the same motion can account for it. Q.E.D.

Several conclusions can be drawn from the above theorems. As long as an underlying motion can account for a four-point mapping, it can account for any extended mapping. If there are different consistent entended mappings, then the structures can be different even for the same motion. If there is a unique motion for a 4-point sub-object, then the same unique motion accounts for all extended consistent mappings obtained from this sub-object. If there is no motion for this hypothetical 4-point object, then there is no motion to account for an extended mapping (if it exists).

3. Conclusion and Discussion

A systematic method is proposed to solve the problems of correspondence, motion, scale and structure. The problems are defined as: Given two consecutive images of an n-point object which undergoes 3D rotation, translation and scaling. Our issues are (i) How can we match the corresponding elements in the two images due to the movement of the object? Can all the possible mappings be found? (ii) What underlying motions and associated depth components of these points could account for the two images? (iii) Can the object be recovered uniquely? This formulation of the problems referred to n-points problem is in the most general setting and does not assume attributes or features.

We discuss whether an n-points problem can be reduced to a set of fewer-points problem. Two aspect of this question arise. The first aspect is the reduction step which enables one to study n-points problem on more manageable sets without affecting the answers. The second aspect is to analyze these problems on small sets. This paper mainly addresses the reduction step along with several observations about small-set problems. A detail solution to small-set problems can be found in [6].

We present a method which reduces an n-point problems to a set of 4-point problems. The effort of reduction takes $O(n)$ steps and it also takes $O(n)$ steps to construct all possible mappings of an n-point sets from the solution to a 4-point problem. One of the conclusions is that observing more than four points in only two views would not help to determine the underlying motions. This conclusion, by contrast, is not true if, instead of parallel projection, perspective projection is used, as was demonstrated in [3]. Other results include (1) Coplanarity condition of four points in two views. (2) How to detect degerate motion. (3) Recovering the tilt direction of the rotational axis using four points in two views. (4) Recovering the scaling factor.

"Four points three views" theorem in [4] can also be addressed by this technique which are currently under preparation.

References

1. Mike Brady (1981), "Special Volume on Computer Vision", *Artifical Intelligence, Volume 17.*

2. Baker (1981), "Depth from Edge and Intensity Based Stereo", Ph.D thesis, AIM-347, Computer Science Department, Stanford University.

3. Tsai, R. Y. and T.S. Huang (1984), "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces", *IEEE Trans. PAMI 6,* 13-27.

4. Ullman, S. (1979), "The Interpretation of Visual Motion", MIT press.

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

6. Lee C.H. (1986), "Matching and Motion of Four Points in Two Views" Technical Report, Department of Computer Science, Purdue University.