International Journal of Computer Vision, 13, 3, 295–309 (1994)

Mirror Uncertainty and Uniqueness Conditions for Determining Shape and Motion from Orthographic Projection

XIAOPING HU AND NARENDRA AHUJA
Coordinated Science Laboratory and Beckman Institute, University of Illinois at Urbana-Champaign, 405 N. Mathews Avenue, Urbana, IL 61801, USA
e-mail: xh, ahuja@vision.csl.uiuc.edu

Received March 1, 1993; accepted November 30, 1993

Abstract. This paper presents new forms of necessary and sufficient conditions for determining shape and motion to within a mirror uncertainty from monocular orthographic projections of any number of point trajectories over any number of views. The new forms of conditions use image data only and can therefore be employed in any practical algorithms for shape and motion estimation. We prove that the mirror uncertainty for the three view problem also exists for a long sequence: if shape **S** is a solution, so is its mirror image **S**′ which is symmetric to **S** about the image plane. The necessary and sufficient conditions for determining the two sets of solutions are associated with the rank of the measurement matrix **W**.

If the rank of **W** is 3, then the original 3D scene points cannot be coplanar and the shape and motion can be determined to within a mirror uncertainty from the image data if and only if there are three distinct views. This condition is different from Ullman's theorem (which states that *three distinct views of four noncoplanar points suffice to determine the shape and motion up to a reflection*) in two aspects: (1) it is expressed in terms of image data; (2) it applies to a long image sequence in a homogeneous way.

If the rank of **W** is 2 and the image points in at least one view are not colinear in the image plane, then there are two possibilities: either the motion is around the optical axis, or the 3-D points all lie on the same plane. In the first case, the motion can be determined uniquely but the shape is not determined. In the second case, a necessary and sufficient condition is to be satisfied and at least 3 point trajectories over at least 3 distinct views are needed to determine the shape in each view to within a mirror uncertainty, and the number of motion solutions is equal to the combinatorial number of the possible positions of the plane in different views. The necessary and sufficient condition is associated with the rank of a matrix **C**: if **C** has a rank of 1, the plane is undetermined; if **C** has a rank of 2 (implying there are exactly 3 distinct views), then a necessary and sufficient condition, whose physical meaning is not completely clear, is to be satisfied to determine the plane to within 2 sets; if **C** has a rank of 3 (implying there are 4 or more distinct views), then the plane can always be determined to within two sets.

If the rank of **W** is 2 or 1 and the image points in each view are colinear in the image plane, then the three dimensional motion problem reduces to a two dimensional motion problem. In this case, the uniqueness condition is associated with the rank of the reduced measurement matrix Ψ. If Ψ has a rank of 2, then the original 3D points cannot be colinear in the space and the shape and motion can be determined to within two sets if and only if three or more views are distinct. If Ψ has a rank of 1, there are two possibilities: if the rows of Ψ are identical, then either the original 3D points are not colinear and the motion is zero, or the points are colinear and possibly move between two mirror symmetric positions; if the rows of Ψ are not identical, then the motion is not determined.

All proofs are constructive and thus define an algorithm for determining the uniqueness of solution as well as for estimating shape and motion from point trajectories.

1 Introduction

The primary focus of the motion researchers has been on the perspective projection. This is probably due to the fact that perspective projection models the imaging process of ordinary cameras more accurately and better conditions the problem of three-dimensional estimation. However, in many situations, orthographic projection is a satisfactory approximation of the imaging process. For example, when a telephoto lens is used, the imaging process can be approximated by orthographic projection provided the motion and size of a moving object in the direction of the optical axis are negligible compared to the average object distance, although a scale constant may be involved (Kanatani 1986). In medical imaging (such as X-ray), the imaging process can be considered as involving orthographic projection. Recently, good experimental results with real image data using orthographic projection models have been reported (Debrunner and Ahuja 1990; 1992a, 1992b; Tomasi and Kanade 1990; 1992). These results show that under certain conditions the orthographic projection is a suitable model for imaging process. Since the three-dimensional estimation problem is much easier for orthographic projection than for perspective projection, the orthographic projection case deserves further investigation.

Ullman (Ullman 1977; 1979) has shown that given three distinct views of four noncoplanar points, the structure of the points can be determined up to a reflection. This condition gives a clear physical meaning and may be generalized to more views. However, a problem with Ullman's condition is that the condition requires a prior knowledge that the views are distinct and the points given are not coplanar. In the process of structure and motion estimation, this knowledge is often unavailable. Therefore, it is desired to have a condition that is expressed in terms of image data only. This motivated some earlier papers (Aloimonos and Brown 1986; Hu and Ahuja 1991a). Huang and Lee (Huang and Lee 1989) also reconfirmed Ullman's results, although there is an additional uncertain situation which cannot be resolved by the methods that enforce motion consistency only (Hu and Ahuja 1991a). Another important question that remains unanswered is whether the mirror uncertainty in the three-view case also exists when more than three views are available. A robust and homogeneous method for determining the uniqueness of the solution for any number of views of an arbitrary surface from the image data has not been available.

In this paper we shall formally prove that for orthographic projection, two mirror symmetric solutions always exist. Consequently, for the structure and motion problem under orthographic projection, uniqueness means two mirror symmetric solutions. We use the term shape instead of structure in this paper because for orthographic projection, the depth information obtained is different from that obtained in perspective projection. In perspective projection, the depths are determined to within a *multiplicative* constant, and hence the object structure and relative distance to the camera coordinate system are determined. In orthographic projection, as will be seen soon, the depths are determined to within an *additive* constant, and hence the object shape and size are determined while the object distance to the camera is not determined.

The central objective of this paper is to obtain necessary and sufficient conditions under which motion and shape are determined to within a minimum number of sets from monocular point trajectories. These conditions apply to any number of point trajectories over any number of views and are expressed in terms of *image data* only. Therefore, they can be actually used in the motion estimation algorithms to determine if a given set of image data admits a unique pair of solutions of shape and motion. For the general case in which the 3D points are noncoplanar, the uniqueness condition is almost equivalent to Ullman's condition (Ullman 1977; 1979) that *at least three views are distinct* (a distinct view means involving a rotation around an axis other than the optical axis). Therefore, the new conditions

The support of National Science Foundation and Defense Advanced Research Projects Agency under grant IRI-89-02728, and the US Army Advanced Construction Technology Center under grant DAAL 03-87-K-0006, is gratefully acknowledged.

are generalized forms of Ullman's condition expressed in terms of image data. The new forms of the conditions are thus more useful in the process of shape and motion estimation, because they can be used to determine not only whether the solution is unique, but also whether the views are distinct and whether the 3D points are coplanar.

In the rest of this paper, all different situations are enumerated and the necessary and sufficient conditions for determining the shape and motion to the minimum number of solutions are established. Section 2 points out a mirror uncertainty that cannot be resolved from the image data and then defines the uniqueness problem for orthographic projection. Section 3 presents the necessary and sufficient condition for the uniqueness problem when the measurement matrix has a rank of 3. In this case, the uniqueness condition is almost equivalent to Ullman's condition. Section 4 first shows that when the measurement matrix has a rank of 2, multiple solutions may result; then necessary and sufficient conditions are obtained for the cases in which rotation is around optical axis or the 3-D points are coplanar. Section 5 presents a necessary and sufficient condition for the case where the image points in each view are all colinear. Section 6 summarizes the paper.

2 The Uniqueness Problem

For the uniqueness problem, we assume that the image data are generated by rigid motions. When we derive the uniqueness conditions, we assume that the data are noiseless. A brief discussion about how to apply the results to noisy situations is also given. The formulation in this paper is based on the factorization method introduced by Debrunner and Ahuja (Debrunner and Ahuja 1990; 1992) for constant motion and Tomasi and Kanade (Tomasi and Kanade 1990; 1992) for general motion. The factorization method enforces both shape and motion consistency. Therefore, the conditions obtained using this method are algorithm-independent since all constraints about rigid motion are utilized.

For orthographic projection, the basic two-view observation model is

$$\Xi_i' = \begin{bmatrix} x_i' \\ y_i' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ Z_i \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{i}^{\mathrm{T}} \\ \mathbf{j}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \Xi_i \\ Z_i \end{bmatrix} + \mathbf{T}_1 = \mathbf{Q}\mathbf{X}_i + \mathbf{T}_1, \qquad (2.1)$$

where $\Xi_i = [x_i, y_i]^{\mathrm{T}}$ is an image point before motion, and $\Xi_i' = [x_i', y_i']^{\mathrm{T}}$ is the correspondence of Ξ_i after motion. Assume there are P pairs of point correspondences with Ξ_C' being the centroid of Ξ_i', $i = 1, 2, \ldots, P$, and X_C being the centroid of $\mathbf{X}_i$, $i = 1, 2, \ldots, P$. It is obvious that Ξ_C' and X_C also satisfy Equation (2.1). Therefore, we can eliminate the *subtranslation vector* $\mathbf{T}_1$ by subtracting Ξ_C' from the left hand side of Equation (2.1) and X_C from the right hand side to get

$$\Xi_i' - \Xi_C' = \mathbf{Q}(\mathbf{X}_i - X_C). \qquad (2.2)$$

In the discussion below, we shall use Ξ_i' to represent $\Xi_i' - \Xi_C'$ and Ξ_i to represent $\Xi_i - \Xi_C$, thus obtaining the following equation

$$\Xi_i' = \mathbf{Q}\mathbf{X}_i. \qquad (2.3)$$

As long as the data is normalized (by subtracting the centroid from each point), the above equation holds. We say that **the data is normalized if from every coordinate for each view the centroid coordinates for that view have been subtracted.** The notation for normalized and unnormalized data will be the same, and unless stated otherwise, image data are assumed to have been normalized for all orthographic projection. However, the 3D points in the necessary and sufficient conditions always mean the points in the original coordinate system, unless stated otherwise.

Now assume we are given P trajectories of points over F frames. Let Ξ_i^f denote the ith point in the fth view, and

$$\mathbf{w}_f = \begin{bmatrix} \Xi_1^f & \cdots & \Xi_P^f \end{bmatrix}. \qquad (2.4)$$

We then have the following equation

$$\mathbf{W}_{2F\times P} = \begin{bmatrix} \mathbf{w}_1 \\ \cdots \\ \mathbf{w}_F \end{bmatrix}_{2F\times P} = \begin{bmatrix} \mathbf{Q}_1 \\ \cdots \\ \mathbf{Q}_F \end{bmatrix}_{2F\times 3} \begin{bmatrix} \mathbf{X}_1^1 & \cdots & \mathbf{X}_P^1 \end{bmatrix}_{3\times P} \equiv \mathbf{MS}, \qquad (2.5)$$

where $\mathbf{X}_i^1$ is the normalized 3-D coordinate of the ith point in the first frame, $\mathbf{Q}_i$ consists of the first two rows of the rotation matrix $\mathbf{R}_i$ between the ith frame and the first frame. Therefore,

$$\mathbf{Q}_f = \begin{bmatrix} \mathbf{i}_f^{\mathrm{T}} \\ \mathbf{j}_f^{\mathrm{T}} \end{bmatrix}, f = 1, \ldots, F, \tag{2.6}$$

with

$$\mathbf{i}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{j}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \tag{2.7}$$

and

$$\mathbf{i}_f \cdot \mathbf{i}_f = 1, \quad \mathbf{j}_f \cdot \mathbf{j}_f = 1, \quad \mathbf{i}_f \cdot \mathbf{j}_f = 0, \quad f = 1, \ldots, F. \tag{2.8}$$

We shall call $\mathbf{W}$ the *measurement matrix*, $\mathbf{w}_f$ the fth entry of $\mathbf{W}$, $\mathbf{M}$ the *motion matrix*, and $\mathbf{S}$ the *shape matrix*. If $\mathbf{MS}$ is a factorization of $\mathbf{W}$ with the rows of $\mathbf{M}$ satisfying Equation (2.8), then $\mathbf{MS}$ is called a *motion factorization* of $\mathbf{W}$.

Although the fth entry $\mathbf{Q}_f$ of $\mathbf{M}$ contains only two rows of the fth rotation matrix $\mathbf{R}_f$, the third row $\mathbf{k}_f^{\mathrm{T}}$ of $\mathbf{R}_f$ is uniquely determined by the first two rows through

$$\mathbf{k}_f = \mathbf{i}_f \times \mathbf{j}_f. \tag{2.9}$$

In other words, the rotation $\mathbf{R}_f$ is given by

$$\mathbf{R}_f = \begin{bmatrix} \mathbf{i}_f^{\mathrm{T}} \\ \mathbf{j}_f^{\mathrm{T}} \\ (\mathbf{i}_f \times \mathbf{j}_f)^{\mathrm{T}} \end{bmatrix}. \tag{2.10}$$

Therefore, if $\mathbf{M}$ and $\mathbf{S}$ can be determined uniquely, the rotation matrices $\mathbf{R}_f$, and the translation vectors $\mathbf{T}_f$, $f = 1, \ldots, F$, can all be determined to within a constant. But the absolute distances and the translations along the optical axis cannot be determined in anyway.

Before we give a formal definition of the uniqueness problem, let us note a *mirror uncertainty*: if $\mathbf{M}$ and $\mathbf{S}$ consist of a motion factorization of the measurement matrix $\mathbf{W}$, so do $\mathbf{MJ}$ and $\mathbf{J}^{-1}\mathbf{S}$, where

$$\mathbf{J} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{2.11}$$

and $\mathbf{J}$ is called the *mirror-reflection matrix*. To show this, we need only to note that if the rows of $\mathbf{M}$ satisfy Equations (2.7) and (2.8), so do the rows of $\mathbf{MJ}$. It is interesting to note that the shape matrix $\mathbf{J}^{-1}\mathbf{S} = \mathbf{JS}$ is just a mirror reflection of $\mathbf{S}$ about the image plane: the sign of the third row of $\mathbf{S}$ is changed. That is, the two shapes are mirror images of each other, symmetric about the image plane. Since the data have been normalized by subtracting the mean from each row of $\mathbf{W}$, the positive depth constraint does not apply here. Consequently, the rotation matrix $\mathbf{R}_f$ and translation vectors $\mathbf{T}_f$, for each $f \geq 2$, will be affected.

This mirror uncertainty has been observed by Ullman (Ullman 1977; 1979) and Huang and Lee (Huang and Lee 1989) for three views, but exists for any number of views and cannot be resolved by the correspondence data. Consequently, the uniqueness problem for estimating shape and motion from monocular orthographic projections is defined as follows.

The Uniqueness Problem of Orthographic Projection: For a given measurement matrix $\mathbf{W}_{2F\times P}$, the shape and motion are said to be determined uniquely if and only if $\mathbf{W}$ can be decomposed, up to a mirror uncertainty, into the product of two matrices $\mathbf{M}_{2F\times 3}$ and $\mathbf{S}_{3\times P}$ such that $\mathbf{M}$ can be represented by

$$\mathbf{M} = \begin{bmatrix} \mathbf{Q}_1 \\ \vdots \\ \mathbf{Q}_F \end{bmatrix} \tag{2.12}$$

with $\mathbf{Q}_f$, $f = 1, \ldots, F$, satisfying Equations (2.6), (2.7) and (2.8). And, the two solutions defined by $\mathbf{M}$, $\mathbf{S}$ and $\mathbf{MJ}$, $\mathbf{JS}$ are called the **mirror conjugate** pair.†

Let us now examine the relationship of the rotation matrices in a mirror conjugate pair. Without loss of generality, let us consider the rotation of the fth frame. We start with $\mathbf{Q}_f$. It is straightforward that if

$$\mathbf{Q}_f = \begin{bmatrix} i_{f1} & i_{f2} & i_{f3} \\ j_{f1} & j_{f2} & j_{f3} \end{bmatrix} \tag{2.13}$$

is the fth entry of $\mathbf{M}$, then the conjugate of $\mathbf{Q}_f$

$$\mathbf{Q}_f^c = \begin{bmatrix} i_{f1} & i_{f2} & -i_{f3} \\ j_{f1} & j_{f2} & -j_{f3} \end{bmatrix} \tag{2.14}$$

is the fth entry of **MJ**. As a result (see Equation (2.10)), if

$$\mathbf{R}_f = \begin{bmatrix} i_{f1} & i_{f2} & i_{f3} \\ j_{f1} & j_{f2} & j_{f3} \\ k_{f1} & k_{f2} & k_{f3} \end{bmatrix} \tag{2.15}$$

is the fth rotation matrix defined by **M**, then the conjugate of $\mathbf{R}_f$

$$\mathbf{R}_f^c = \begin{bmatrix} i_{f1} & i_{f2} & -i_{f3} \\ j_{f1} & j_{f2} & -j_{f3} \\ -k_{f1} & -k_{f2} & k_{f3} \end{bmatrix}, \tag{2.16}$$

is the fth rotation matrix defined by **MJ**. Using the three-angle representation of a rotation matrix (Ullman 1979), we conclude that if

$$\mathbf{R}_f = \mathbf{A}_X(\theta_X^f)\mathbf{A}_Y(\theta_Y^f)\mathbf{A}_Z(\theta_Z^f), \tag{2.17}$$

where $\mathbf{A}_X$, $\mathbf{A}_Y$, and $\mathbf{A}_Z$ are rotations around X, Y, and Z axis respectively, then

$$\mathbf{R}_f^c = \mathbf{A}_X(-\theta_X^f)\mathbf{A}_Y(-\theta_Y^f)\mathbf{A}_Z(\theta_Z^f). \tag{2.18}$$

That is, the rotation angles θ_X^f and θ_Y^f have opposite signs in the conjugate pair of rotation matrices.

Equations (2.6), (2.7), and (2.8) are the necessary and sufficient conditions for a matrix **M** of the form (2.12) to be a motion matrix. The motion problem for orthographic projection is then reduced to recovering the motion matrix and the shape matrix from a given measurement matrix.

First let us note the *rank principle* (Tomasi and Kanade 1990, 1992; Hu and Ahuja 1991a). It is obvious that the ranks of **M**, **S**, and **W** are all at most 3. Therefore, we have the following fact:

FACT 2.1. The motion matrix **M** and the shape matrix **S** all have a rank of at most 3. For noiseless data, the measurement matrix **W** also has a rank of at most 3. If 3 or more trajectories over 2 or more frames are available, then all three matrices have a rank of at least 1.†

This rank principle is the basis of the factorization method that has been developed by Debrunner and Ahuja and Tomasi and Kanade. In the rest of this paper, we shall start with Tomasi and Kanade's factorization method and then obtain the conditions for unique solution of the motion and shape. The difference between the method below and Tomasi & Kanade's algorithm is that our method applies also to degenerate cases and replaces the quadratic fitting method in Tomasi & Kanade's algorithm by factorization method.

Obviously, at least 3 point trajectories over at least 2 frames are needed to determine the motion and/or the shape uniquely. Hereafter we shall assume that at least 3 trajectories over at least 2 frames are available. Otherwise, the motion problem is not determined.

Even when the rotation is determined, there is still a constant in the depths and the subtranslation vectors which cannot be determined in the original coordinate system. Increasing the number of views or correspondences does not remove this uncertainty. But overall, there is only one unknown constant. Let us consider only the case of increasing the number of views here. Let $\mathbf{T}_1^f$ and $\mathbf{R}_f$ be the subtranslation vector and the rotation between the fth and the first views. Consider the motion of point Ξ_C. Let its position in the fth view be Ξ_C^f with depth Z_C^f. From Equation (2.1) we have

$$\begin{bmatrix} r_{13}^f \\ r_{23}^f \end{bmatrix} Z_C^1 + \mathbf{T}_1^f = \Xi_C' - \begin{bmatrix} r_{11}^f & r_{12}^f \\ r_{21}^f & r_{22}^f \end{bmatrix} \Xi_C = \zeta_C^f,$$
$$f = 2, \ldots, F. \tag{2.19}$$

The above equation can be written into the vector form

$$\mathbf{T} = \mathbf{r} Z_C^1 + \zeta, \tag{2.20}$$

where

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_1^2 \\ \vdots \\ \mathbf{T}_1^F \end{bmatrix}, \quad \mathbf{r} = -\begin{bmatrix} r_{13}^2 \\ r_{23}^2 \\ \vdots \\ r_{13}^F \\ r_{13}^F \end{bmatrix}, \quad \zeta = \begin{bmatrix} \zeta_C^2 \\ \vdots \\ \zeta_C^F \end{bmatrix}, \tag{2.21}$$

and the superscripts do not mean *power*. For a given Z_C^1, the translation can be determined by Equation (2.20). However, no restriction is imposed on Z_C except that one can require it to be positive. More correspondences will give more equations similar to (2.20) which will not help to remove the uncertainty since $\mathbf{r}$ is the same for all points.

Quite similarly, another constant is involved in the translation components along the optical axis. Let t_3^f be the third component of the translation between the fth and the first views. Then from the motion equation (Equation (2) in [6]) we have

$$Z_C^f = \begin{bmatrix} r_{31}^f & r_{32}^f \end{bmatrix} \Xi_C^1 + r_{33}^f Z_C^1 + t_3^f. \tag{2.22}$$

For a given Z_C^1, there is another constant in Z_C^f, and t_3^f, $f = 2, \ldots, F$, which cannot be determined.

Thus we have the following fact:

FACT 2.2. For a given solution of the rotations, the subtranslations and the depths can be determined to within to a constant. An additional constant is involved in the depths and the third components of the translations which also cannot be determined. The positive depth constraint does not help remove the uncertainties in the translations and depths and does not help remove the mirror uncertainty of the rotations.†

With this fact in mind, we will concentrate on the solution of the rotations and shape vectors.

3 When W Has a Rank of 3

The most general situation occurs when $\mathbf{W}$ has a rank of 3. We first have the following theorem which is expressed in terms of the correspondence data.

THEOREM 3.1. *If the $2F \times P$ measurement matrix $\mathbf{W}$ has a rank of 3, $\mathbf{W}$ can be factorized as*

$$\mathbf{W} = \mathbf{L}_{2F\times 3}\mathbf{Y}_{3\times P} = \begin{bmatrix} \mathbf{u}_1^{\mathrm{T}} \\ \mathbf{v}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{u}_F^{\mathrm{T}} \\ \mathbf{v}_F^{\mathrm{T}} \end{bmatrix} [\mathbf{y}_1 \quad \cdots \quad \mathbf{y}_P], \tag{3.1}$$

where both $\mathbf{L}$ and $\mathbf{Y}$ have a rank of 3. Then if and only if matrix

$$\mathbf{D} = \begin{bmatrix} u_{11}^2 & 2u_{11}u_{12} & 2u_{11}u_{13} & u_{12}^2 & 2u_{12}u_{13} & u_{13}^2 \\ v_{11}^2 & 2v_{11}v_{12} & 2v_{11}v_{13} & v_{12}^2 & 2v_{12}v_{13} & v_{13}^2 \\ u_{11}v_{11} & u_{11}v_{12}+v_{11}u_{12} & u_{11}v_{13}+v_{11}u_{13} & u_{12}v_{12} & u_{12}v_{13}+v_{12}u_{13} & v_{13}u_{13} \\ & & \cdots & \cdots & & \\ u_{F1}^2 & 2u_{F1}u_{F2} & 2u_{F1}u_{F3} & u_{F2}^2 & 2u_{F2}u_{F3} & u_{F3}^2 \\ v_{F1}^2 & 2v_{F1}v_{F2} & 2v_{F1}v_{F3} & v_{F2}^2 & 2v_{F2}v_{F3} & v_{F3}^2 \\ u_{F1}v_{F1} & u_{F1}v_{F2}+v_{F1}u_{F2} & u_{F1}v_{F3}+v_{F1}u_{F3} & u_{F2}v_{F2} & u_{F2}v_{F3}+v_{F2}u_{F3} & u_{F3}v_{F3} \end{bmatrix}_{3F\times 6} \tag{3.2}$$

has a rank of 6, the motion and shape are both determined to within a mirror uncertainty, where u_{ij} and v_{ij} are the elements of $\mathbf{u}_i$ and $\mathbf{v}_i$ respectively, i.e.,

$$\mathbf{u}_f^{\mathrm{T}} = [u_{f1} \quad u_{f2} \quad u_{f3}],$$
$$\mathbf{v}_f^{\mathrm{T}} = [v_{f1} \quad v_{f2} \quad v_{f3}], \quad f = 1, \ldots, F. \tag{3.3}$$

Proof. That there exists a factorization of the form of Equation (3.1) for $\mathbf{W}$ is a result of linear algebra and is obvious from the motion equation (2.5). The only thing we need to prove is that if and only if $\mathbf{D}$ has a rank of 6, then the motion and shape are both determined to within a mirror uncertainty. First let us note that if $\mathbf{L}$ and $\mathbf{Y}$ are a factorization of $\mathbf{W}$, so are $\mathbf{LA}$ and $\mathbf{A}^{-1}\mathbf{Y}$ for any 3×3 invertible matrix $\mathbf{A}$. However, only when $\mathbf{LA}$ can be represented in the form of Equation (2.12) with Equations (2.6), (2.7) and (2.8) satisfied, $\mathbf{LA}$ and $\mathbf{A}^{-1}\mathbf{Y}$ consist of a valid motion and shape factorization.

Now let us first prove that if $\mathbf{L}_{2F\times 3}$, $\mathbf{Y}_{3\times P}$ and $\mathbf{M}_{2F\times 3}$, $\mathbf{S}_{3\times P}$ are two factorizations of $\mathbf{W}$ and all matrices have a rank of 3, then the two $2F \times 3$ matrices $\mathbf{L}$ and $\mathbf{M}$ can be related by the following form

$$\mathbf{M} = \mathbf{LA} \tag{3.4}$$

with $\mathbf{A}$ a 3×3 invertible matrix. Since

$$\mathbf{W} = \mathbf{LY} = \mathbf{MS}, \tag{3.5}$$

postmultiplying the above equation with $\mathbf{Y}^T$ leads to

$$\mathbf{W}\mathbf{Y}^T = \mathbf{L}(\mathbf{Y}\mathbf{Y}^T) = \mathbf{M}(\mathbf{S}\mathbf{Y}^T). \tag{3.6}$$

Since $\mathbf{Y}$ and $\mathbf{S}$ all have a rank of 3, $\mathbf{Y}\mathbf{Y}^T$ and $\mathbf{S}\mathbf{Y}^T$ are invertible. Therefore $\mathbf{M}$ and $\mathbf{L}$ are related through

$$\mathbf{M} = \mathbf{L}(\mathbf{Y}\mathbf{Y}^T)(\mathbf{S}\mathbf{Y}^T)^{-1}. \tag{3.7}$$

Consequently we need only to find a 3×3 invertible matrix $\mathbf{A}$ such that $\mathbf{LA}$ is a motion matrix. If $\mathbf{A}$ is determined uniquely, so are the motion and shape. If $\mathbf{A}$ is not determined uniquely, neither are the motion and shape. To determine $\mathbf{A}$ uniquely, Equation (2.8) is used to enforce that $\mathbf{LA}$ is a motion matrix. Therefore we have the following equations:

$$\begin{aligned} &\mathbf{u}_f^T\mathbf{A}\mathbf{A}^T\mathbf{u}_f = 1, \quad \mathbf{v}_f^T\mathbf{A}\mathbf{A}^T\mathbf{v}_f = 1, \\ &\mathbf{u}_f^T\mathbf{A}\mathbf{A}^T\mathbf{v}_f = 0, \quad f = 1, \ldots, F. \end{aligned} \tag{3.8}$$

Let

$$\mathbf{P} = \mathbf{A}\mathbf{A}^T = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_2 & p_4 & p_5 \\ p_3 & p_5 & p_6 \end{bmatrix}, \tag{3.9}$$

and

$$\begin{aligned} \mathbf{u}_f^T &= [u_{f1} \quad u_{f2} \quad u_{f3}], \\ \mathbf{v}_f^T &= [v_{f1} \quad v_{f2} \quad v_{f3}], \end{aligned} \tag{3.10}$$

then we can get the following equation for $\mathbf{P}$:

$$\mathbf{D}_{3F\times 6}e(\mathbf{P})_{6\times 1} = \mathbf{B}_{3F\times 1}, \tag{3.11}$$

where $\mathbf{D}$ is represented in Equation (3.2) and

$$e(\mathbf{P}) = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 1 \\ 1 \\ 0 \end{bmatrix}_{3F\times 1}. \tag{3.12}$$

Therefore, if and only if $\mathbf{D}$ has a rank of 6, $\mathbf{P}$ can be uniquely determined from Equation (3.11). Once $\mathbf{P}$ is determined, $\mathbf{A}$ can be determined up to an orthonormal matrix by factorizing a symmetric $\mathbf{P}$ into $\mathbf{A}\mathbf{A}^T$ (the factorization of $\mathbf{P}$ is discussed in detail in the appendix). That is, if $\mathbf{A}$ is a symmetric factorization of $\mathbf{P}$, so is $\mathbf{AU}$ for any orthonormal matrix $\mathbf{U}$. Assume $\mathbf{A}$ is one symmetric factorization of $\mathbf{P}$ obtained, e.g., by the singular value decomposition (SVD) technique, and $\mathbf{AU}$ is the correct factorization of $\mathbf{P}$. To determine $\mathbf{U}$, noticing that the rotation of the first frame is known, we require

$$\begin{gathered} \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{v}_1^T \end{bmatrix} \mathbf{AU} = \mathbf{Q}_1, \quad \text{or} \\ \mathbf{U}^T(\mathbf{A}^T[\mathbf{u}_1 \quad \mathbf{v}_1]) = [\mathbf{i}_1 \quad \mathbf{j}_1]. \end{gathered} \tag{3.13}$$

$\mathbf{U}$ can be determined up to a mirror reflection from

$$\begin{aligned} \mathbf{U}_{1,2}^T &= [\mathbf{i}_1 \quad \mathbf{j}_1 \quad \mathbf{i}_1 \times \mathbf{j}_1] \\ &\quad \times \left[\mathbf{A}^T\mathbf{u}_1 \quad \mathbf{A}^T\mathbf{v}_1 \quad \pm(\mathbf{A}^T\mathbf{u}_1) \times (\mathbf{A}^T\mathbf{v}_1)\right]^{-1} \\ &= \left[\mathbf{A}^T\mathbf{u}_1 \quad \mathbf{A}^T\mathbf{v}_1 \quad \pm(\mathbf{A}^T\mathbf{u}_1) \times (\mathbf{A}^T\mathbf{v}_1)\right]^{-1}, \end{aligned} \tag{3.14}$$

where we have used the knowledge that $[\mathbf{i}_1\ \mathbf{j}_1\ \mathbf{i}_1 \times \mathbf{j}_1]$ is identity matrix. The $\pm$ sign corresponds to the mirror uncertainty discussed in Section 2. Since for noiseless data, there always exists at least one solution, the inverse of the above equation must exist.

We have thus proved the theorem. Q.E.D.

The above theorem provides not only a condition but also a method for determining the uniqueness of the solution for a given set image trajectories. Several questions may arise for the above method.

The first question is whether the solution of $\mathbf{P}$ from Equation (3.12) is always positive semi-definite. If not, then $\mathbf{P}$ may not find a factorization into $\mathbf{A}\mathbf{A}^T$ with $\mathbf{A}$ a real matrix. For noiseless data, if $\mathbf{P}$ is uniquely determined, then $\mathbf{P}$ must be positive semi-definite; otherwise, $\mathbf{P}$ will not have a real symmetric decomposition and hence $\mathbf{W}$ will not have a motion factorization, contradicting the assumption that the data are generated by real, rigid motions. For noisy data, the positive semi-definite property of $\mathbf{P}$ may not be warranted by the solution from Equation (3.12). However, in our hundreds of simulations with reasonable amount of noise (10% or less), we have never encountered a single case in which $\mathbf{P}$ is not positive semi-definite.

Even when it indeed occurs that $\mathbf{P}$ is not positive semi-definite, one can search for a real matrix $\mathbf{A}$, which minimizes $\|\mathbf{P} - \mathbf{A}\mathbf{A}^T\|$ and replace $\mathbf{P}$ by $\mathbf{A}\mathbf{A}^T$, as was done by Tomasi and Kanade (Tomasi and Kanade 1992). Then, the rest of the method in the theorem can still be used.

The second question is what the rank condition means or in what physical situations the $\mathbf{D}$ matrix has a full rank. First, since $\mathbf{W}$ has a rank of 3, the shape matrix $\mathbf{S}$ must also have a rank of 3. $\mathbf{S}$ has a rank of 3 if and only if the scene points used are not coplanar in the original coordinate system. Then according to Ullman's classic theorem (Ullman 1977, 1979), three distinct views (a rotation around an axis other than the optical axis must be involved between any two distinct views) suffice to determine the motion up to a reflection. On the other hand, if the whole sequence contains exactly two distinct views (in this case, $\mathbf{W}$ may still have a rank of 3 but $\mathbf{D}$ cannot have a rank of 6), the shape and motion are not determined. If the whole sequence contains only one distinct view (all rotations are around the optical axis), then $\mathbf{M}$ and hence $\mathbf{W}$ have a rank of 2. This situation is discussed in the next section. In other words, when $\mathbf{W}$ has a rank of 3, then $\mathbf{D}$ has a full column rank of 6 if and only if there are 3 distinct views. A direct rank checking procedure (cf. Section 5 for a special case) can be used to verify this fact, which is very long and does not provide more insight into the problem. Therefore, we refer to Ullman's theorem for an indirect proof. Then, what is gained in Theorem 3.1 as it is almost equivalent to Ullman's theorem? There are two major distinctions between Theorem 3.1 and Ullman's theorem. First, Theorem 3.1 is expressed in terms of image data while Ullman's theorem is expressed in terms of the 3D structure of the points and the number of distinct views. For a given set of image data, the number of distinct views and the surface structure of the points cannot be determined before solving the shape and motion. Therefore, Ullman's theorem can hardly be applied in the process of shape and motion estimation while Theorem 3.1 can be used in any practical algorithms to determine the uniqueness of the solution. Second, Ullman's theorem deals with three views while Theorem 3.1 applies to any number of views in a homogeneous way. For a given sequence of images, it may occur that some views are distinct and some other views are not distinct. It would not be efficient and robust if triples of images are examined each time if the sequence is very long. Theorem 3.1 provides such a homogeneous method for determining the uniqueness of the solution.

The third question is how to apply the rank condition for noisy data. In general, noise will not make a determined problem undetermined. Therefore, for the given image trajectories, if $\mathbf{D}$ matrix has a full rank for the noiseless situation, then the presence of small noise in the data will in general not reduce the rank of $\mathbf{D}$. Therefore, in general, problems exist only in degenerate cases in which either the points are coplanar or there are only two or one distinct views. In either case, noise may increase the rank of $\mathbf{W}$ or $\mathbf{D}$. To overcome this problem, the rank of $\mathbf{W}$ can be replaced by the number of nonzero singular values and the rank of $\mathbf{D}$ can be replaced by the number of nonzero eigenvalues. By examining the number of singular values (or eigenvalues) of $\mathbf{W}$ (or $\mathbf{D}$) that are significantly larger than 0, one can robustly determine the rank of $\mathbf{W}$ (or $\mathbf{D}$) and hence the uniqueness of the solution. Such rank checking method has been actually used in most commercial mathematical software packages (e.g., IMSL 1987). Therefore, the above theorem provides a robust method for determining the uniqueness of the solution.

When the rank of $\mathbf{W}$ is less than 3, the above method cannot be used. This situation is called the *degenerate situation* and is discussed in the following sections. The discussion about the first and the third questions above all apply to the methods to be discussed below. Therefore, we will not repeat them in the rest of the paper.

4 When rank(W) = 2 and rank($\mathbf{w}_i$) = 2 for at Least One i

Now let us discuss the situation where the rank of $\mathbf{W}$ is 2 and $\mathbf{w}_i$ has a rank of 2 for at least one i. The latter requirement is equivalent to that the projections of the points in at least one view are not colinear in the image plane. There are three ways in which **W** can have a rank of 2: either the motion matrix **M** has a rank of 2, or the shape matrix **S** has a rank of 2, or both. We mainly discuss degeneracy caused by shape since a degenerate case caused by motion can be dealt with as if the degeneracy were caused by shape, as will be seen shortly.

Clearly, **S** has a rank of 2 if and only if the original 3-D points lie on a plane. Assume the plane in the first view has a normal $\mathbf{N} = [n_1, n_2, n_3]^T$. A normalized space point $\mathbf{X}_i$ in the first frame satisfies the following equation

$$\mathbf{N}^T\mathbf{X}_i = 0. \tag{4.1}$$

We consider only the general situation where $n_3 \neq 0$. The case $n_3 = 0$ is discussed in the next section. When $n_3 \neq 0$, we can solve for the depth from

$$Z_i = \begin{bmatrix} -\dfrac{n_1}{n_3} & -\dfrac{n_2}{n_3} \end{bmatrix} \Xi_i \equiv [l_1 \quad l_2]\,\Xi_i. \tag{4.2}$$

Then, from the two-view motion equation (2.3) we have

$$\begin{aligned} \Xi_i' = \mathbf{Q}\mathbf{X}_i = \mathbf{Q}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ l_1 & l_2 \end{bmatrix} \Xi_i \\ \equiv \mathbf{Q}\mathbf{L}\Xi_i \equiv \mathbf{A}\Xi_i. \end{aligned} \tag{4.3}$$

Clearly matrix $\mathbf{A} = \mathbf{QL}$ is a 2×2 matrix since

$$A = \mathbf{QL} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ l_1 & l_2 \end{bmatrix}. \tag{4.4}$$

Therefore, when the points are coplanar there exists a 2×2 matrix **A** that relates a point Ξ_i in one image to its correspondence Ξ_i' in the other image.

A correspondence pair Ξ_i' and Ξ_i also satisfy the following equation (Huang and Lee 1989; Hu and Ahuja 1991):

$$[r_{23} \quad -r_{13} \quad r_{32} \quad -r_{31}] \begin{bmatrix} \Xi_i' \\ \Xi_i \end{bmatrix} = 0. \tag{4.5}$$

Using Equation (4.3), Equation (4.5) reduces to the following equation

$$(\mathbf{r}_1^T\mathbf{A} + \mathbf{r}_2^T)\Xi_i = 0, \tag{4.6}$$

where

$$\mathbf{r}_1^T = [r_{23}, \ -r_{13}], \quad \mathbf{r}_2^T = [r_{32}, \ -r_{31}]. \tag{4.7}$$

With three matches of noncolinear points, we can uniquely solve for **A** linearly from (4.3) and get the following equation for r_{ij}, $ij \in \kappa = \{13, 23, 31, 32\}$:

$$\mathbf{r}_1^T\mathbf{A} + \mathbf{r}_2^T = 0, \quad \text{or} \quad \mathbf{r}_2 = -\mathbf{A}_1^T\mathbf{r}_1. \tag{4.8}$$

With the identity

$$\|\mathbf{r}_1\|^2 = r_{13}^2 + r_{23}^2 = \|\mathbf{r}_2\|^2 = r_{31}^2 + r_{32}^2, \tag{4.9}$$

we thus have

$$\mathbf{r}_2^T\mathbf{r}_2 = \mathbf{r}_1^T\mathbf{r}_1 = \mathbf{r}_1^T\mathbf{A}\mathbf{A}_1^T\mathbf{r}_1. \tag{4.10}$$

Let **U** be the orthonormal matrix such that

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}^T\Lambda\mathbf{U}, \tag{4.11}$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2)$ is a diagonal matrix with $\lambda_1 \geq \lambda_2$. Then let

$$\mathbf{x} = \mathbf{U}\mathbf{r}_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{r}_1 = \mathbf{U}^T\mathbf{x}, \tag{4.12}$$

we have

$$\mathbf{r}_1^T\mathbf{r}_1 = \mathbf{x}^T\mathbf{U}\mathbf{U}^T\mathbf{x} = \mathbf{x}^T\mathbf{x} = \mathbf{r}_1^T\mathbf{U}^T\Lambda\mathbf{U}\mathbf{r}_1 = \mathbf{x}^T\Lambda\mathbf{x}, \tag{4.13}$$

wherefrom we get

$$(\lambda_1 - 1)x_1^2 = (1 - \lambda_2)x_2^2 \tag{4.14}$$

after some simplification. To have nontrivial real solution for **x**, we must therefore have

$$\lambda_2 \leq 1 \leq \lambda_1. \tag{4.15}$$

When the above inequality is satisfied, any **x** that satisfies (4.14) can be represented as

$$\mathbf{x}^T = [x_1 \quad x_2] = \alpha\left[\sqrt{1-\lambda_2} \quad \pm\sqrt{\lambda_1 - 1}\right], \tag{4.16}$$

where α is a constant. Using Equations (4.12) and (4.8), we can get the following solution for $\mathbf{r}_1$ and $\mathbf{r}_2$

$$\begin{aligned} \mathbf{r}_1 &= \mathbf{U}^T\mathbf{x} = \alpha\mathbf{U}^T\begin{bmatrix} \sqrt{1-\lambda_2} \\ \pm\sqrt{\lambda_1 - 1} \end{bmatrix}, \\ \mathbf{r}_2 &= -\alpha\mathbf{A}^T\mathbf{U}^T\begin{bmatrix} \sqrt{1-\lambda_2} \\ \pm\sqrt{\lambda_1 - 1} \end{bmatrix}. \end{aligned} \tag{4.17}$$

A special case that deserves particular attention is when

$$\lambda_1 = \lambda_2 = 1. \tag{4.18}$$

In this case, we have

$$\mathbf{r}_1 = \mathbf{r}_2 = 0, \tag{4.19}$$

and the rotation matrix is uniquely determined as

$$\mathbf{R} = \begin{bmatrix} \mathbf{A} & & \begin{matrix} 0 \\ 0 \end{matrix} \\ 0 & 0 & |\mathbf{A}| \end{bmatrix}, \tag{4.20}$$

where $|\mathbf{A}|$ is the determinant of $\mathbf{A}$ which is either 1 or -1. The surface shape is not determined in this case since the depths are not related to the image data in any way. This corresponds to the case when the motion matrix has a rank of 2. That is, degeneracy caused by motion can be dealt with in the same way (cf. Hu and Ahuja 1991) as above. Therefore, if $\mathbf{W}$ has a rank of 2, we can first assume that the degeneracy is caused by shape and solve for $\mathbf{A}$ first. If the eigenvalues of $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ are all 1, then it can be concluded that the degeneracy is caused by motion (the shape is unknown); otherwise, it can be concluded that the degeneracy is caused by shape.

Here we see that when the surface is a plane, there are two sets of solution for r_{ij}, $ij \in \kappa$, that determine r_{ij} to within a scalar. The sign uncertainty in Equation (4.17) cannot be completely removed with correspondence data, as shown indirectly in the following.

Huang and Lee (Huang and Lee 1989) have shown that three views yield at most 16 motion solutions for three views and 4 shape solutions corresponding to the first view, including mirror symmetric solutions. Adding more frames will generally reduce the number of shape solutions for each view and the motion solutions between any two views. However, the method that solves motion first and then shape later is not good for this situation since there are too many motion solutions. We now present a method that solves for the shape first and hence an optimal solution for noisy data can be obtained. With this method we also obtain the uniqueness conditions.

Assume $\mathbf{w}_1$ has a rank of 2 (if $\mathbf{w}_1$ has a rank of 1, but $\mathbf{w}_i$ has a rank of 2 for some i, then exchange the position of $\mathbf{w}_1$ with $\mathbf{w}_i$). Let the plane equation for the first view be (4.2). Then, for each $\mathbf{w}_f$, there exists a 2×2 matrix $\mathbf{A}^f$ such that

$$\mathbf{w}_f = \mathbf{A}^f \mathbf{w}_1 = \begin{bmatrix} \mathbf{a}_1^f & \mathbf{a}_2^f \end{bmatrix} \mathbf{w}_1, \quad f = 2, \ldots, F. \tag{4.21}$$

Because $\mathbf{w}_1$ has a rank of 2, $\mathbf{A}^f$ can be determined uniquely from Equation (4.21). From Equation (4.4) we have

$$\mathbf{A}^f = \mathbf{Q}_f \mathbf{L} = \begin{bmatrix} r_{11}^f & r_{12}^f & r_{13}^f \\ r_{21}^f & r_{22}^f & r_{23}^f \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ l_1 & l_2 \end{bmatrix}, \tag{4.22}$$

where $\mathbf{Q}_f$ is the first two rows of the fth rotation matrix $\mathbf{R}_f$. For planar surface situation, the above equation contains all information available about the rotation matrix $\mathbf{R}_f$. Let

$$\begin{aligned} b_1^f &= \begin{bmatrix} r_{31}^f & r_{32}^f & r_{33}^f \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ l_1 \end{bmatrix}, \\ b_2^f &= \begin{bmatrix} r_{31}^f & r_{32}^f & r_{33}^f \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ l_2 \end{bmatrix}. \end{aligned} \tag{4.23}$$

Then we have

$$\begin{aligned} \begin{bmatrix} \mathbf{a}_1^f & \mathbf{a}_2^f \\ b_1^f & b_2^f \end{bmatrix} &= \begin{bmatrix} r_{11}^f & r_{12}^f & r_{13}^f \\ r_{21}^f & r_{22}^f & r_{23}^f \\ r_{31}^f & r_{32}^f & r_{33}^f \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ l_1 & l_2 \end{bmatrix} \\ &= \mathbf{R}_f \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ l_1 & l_2 \end{bmatrix}. \end{aligned} \tag{4.24}$$

Using the orthonormality of $\mathbf{R}_f$ we can solve for $b_i^f, i = 1, 2$, as

$$b_i^f = \pm\sqrt{1 + l_i^2 - \|\mathbf{a}_i^f\|^2}, \quad i = 1, 2. \tag{4.25}$$

Rotation matrix $\mathbf{R}_f$ can have a solution from Equation (4.24) if and only if (see Hu and Ahuja 1991b)

$$\begin{bmatrix} \mathbf{a}_1^f \\ b_1^f \end{bmatrix} \cdot \begin{bmatrix} \mathbf{a}_2^f \\ b_2^f \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ l_1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ l_2 \end{bmatrix}. \tag{4.26}$$

If the above equation is satisfied, $\mathbf{R}_f$ can be solved for from

$$\mathbf{R}_f = \left[\mathbf{c}_1^f \quad \mathbf{c}_2^f \quad \mathbf{c}_1^f \times \mathbf{c}_2^f\right]\left[\mathbf{d}_1 \quad \mathbf{d}_2 \quad \mathbf{d}_1 \times \mathbf{d}_2\right]^{-1}, \tag{4.27}$$

where

$$\mathbf{c}_1^f = \begin{bmatrix} \mathbf{a}_1^f \\ b_1^f \end{bmatrix}, \mathbf{c}_2^f = \begin{bmatrix} \mathbf{a}_2^f \\ b_2^f \end{bmatrix}, \mathbf{d}_1 = \begin{bmatrix} 1 \\ 0 \\ l_1 \end{bmatrix}, \mathbf{d}_2 = \begin{bmatrix} 0 \\ 1 \\ l_2 \end{bmatrix}. \tag{4.28}$$

Equation (4.26) can be simplified to

$$b_1^f b_2^f = l_1 l_2 - \mathbf{a}_1^f \cdot \mathbf{a}_2^f, \tag{4.29}$$

or

$$\pm\sqrt{1 + l_1^2 - \|\mathbf{a}_1^f\|^2}\sqrt{1 + l_2^2 - \|\mathbf{a}_2^f\|^2} = l_1 l_2 - \mathbf{a}_1^f \cdot \mathbf{a}_2^f. \tag{4.30}$$

From this equation we can see that two motion solutions are associated with each plane solution since the sign of $b_1^f b_2^f$ must be the same as the sign of $l_1 l_2 - \mathbf{a}_1^f \cdot \mathbf{a}_2^f$. Squaring both sides of the above equation and simplifying the result, we get

$$\alpha_f l_1^2 + \beta_f l_1 l_2 + \gamma_f l_2^2 = \delta_f, \; f = 2, \ldots, F, \tag{4.31}$$

where

$$\begin{aligned} \alpha_f &= 1 - \|\mathbf{a}_2^f\|^2, \quad \beta_f = 2\left(\mathbf{a}_1^f \cdot \mathbf{a}_2^f\right), \\ \gamma_f &= 1 - \|\mathbf{a}_1^f\|^2, \quad \delta_f = \left(\mathbf{a}_1^f \cdot \mathbf{a}_2^f\right)^2 - \alpha_f \gamma_f. \end{aligned} \tag{4.32}$$

Equation (4.31) can be used to determine l_1 and l_2. Let

$$\mathbf{C} = \begin{bmatrix} \alpha_2 & \beta_2 & \gamma_2 \\ & \vdots & \\ \alpha_F & \beta_F & \gamma_F \end{bmatrix}_{(F-1)\times 3}, \quad \Delta = \begin{bmatrix} \delta_2 \\ \vdots \\ \delta_F \end{bmatrix}, \tag{4.33}$$

Equation (4.31) becomes

$$\mathbf{C} \begin{bmatrix} l_1^2 \\ l_1 l_2 \\ l_2^2 \end{bmatrix} = \Delta. \tag{4.34}$$

If $\mathbf{C}$ has a rank of 3 (implying $F \geq 4$), then l_1 and l_2 can be determined to within two sets linearly from the above equation. If $\mathbf{C}$ has a rank of 2, then l_1 and l_2 can be determined to within four sets from the above equation. If $\mathbf{C}$ has a rank of 1, then the shape and motion are not determined.

When $\mathbf{C}$ has a rank of 2, the following method can be used to obtain a closed form solution of l_1 and l_2. Without loss of generality. Let us consider that Equation (4.31) yields two independent equations for $f = i$ and $f = j$. Let

$$l_1 = \rho \sin\varphi, \; l_2 = \rho\cos\varphi, \tag{4.35}$$

we then have

$$\rho^2\left(\alpha_f \sin^2\varphi + \beta_f \sin\varphi\cos\varphi + \gamma_f \cos^2\varphi\right) = \delta_f, \quad f = i, j. \tag{4.36}$$

Eliminating ρ from the above two equations and simplifying the resulting equation we can get

$$a\tan^2\varphi + b\tan\varphi + c = 0, \tag{4.37}$$

where

$$\begin{aligned} &a = \alpha_i\delta_j - \alpha_j\delta_i, \quad b = \beta_i\delta_j - \beta_j\delta_i, \\ &c = \gamma_i\delta_j - \gamma_i\delta_j \end{aligned} \tag{4.38}$$

It is clear that $\tan\varphi$ can have two solutions from Equation (4.37) and then ρ can have four solutions from Equation (4.36). In total l_1 and l_2 can be determined to within four sets. The necessary and sufficient condition for determining l_1 and l_2 to within 2 sets is

$$b^2 - 4ac = 0. \tag{4.39}$$

When $\mathbf{C}$ has a rank of 3, l_1 and l_2 can always be determined to within 2 sets linearly. However, for noisy data, an optimal solution is desired. Again, representation (4.35) can be used to obtain a nonlinear least squares solution which needs to search for φ only. Let

$$h_f(\varphi) = \alpha_f \sin^2\varphi + \beta_f \sin\varphi\cos\varphi + \gamma_f\cos^2\varphi, f = 2, \ldots, F. \tag{4.40}$$

Then Equation (4.36) becomes

$$h_f(\varphi)\rho^2 = \delta_f, \quad f = 2, \ldots, F. \tag{4.41}$$

For a given φ, the optimal solution of ρ^2 that minimizes the residue

$$\epsilon^2 \equiv \sum_{f=2}^{F} \left(h_f(\varphi)\rho^2 - \delta_f\right)^2 \tag{4.42}$$

is

$$\rho^2(\varphi) = \frac{\sum_f \delta h_f(\varphi)}{\sum_f h_f^2(\varphi)}. \tag{4.43}$$

With ρ_2 chosen above, the residue becomes, after some simplification,

$$\epsilon^2(\varphi) = \sum_f \delta_f^2 - \frac{\left(\sum_f \delta_f h_f(\varphi)\right)^2}{\sum_f h_f^2(\varphi)}. \tag{4.44}$$

Therefore, the optimal choice of φ is that which maximizes

$$\varpi = \frac{\left(\sum_f \delta_f h_f(\varphi)\right)^2}{\sum_f h_f^2(\varphi)}. \tag{4.45}$$

When $\mathbf{C}$ has a rank of 3, there is only one φ within $[0, \pi)$ that maximizes ϖ. An optimal search is needed to solve for φ and afterwards ρ^2 can be obtained in closed form from Equation (4.43).

We summarize the discussion above as the following theorem.

THEOREM 4.1. *Let the $2F \times P$ measurement matrix $\mathbf{W}$ be written as*

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_F \end{bmatrix}, \tag{4.46}$$

where $\mathbf{w}_f$ is a $2 \times P$ submatrix of the measurement of the fth frame, for $f = 1, 2, \ldots, F$. When $\mathbf{W}$ and $\mathbf{w}_i$ for some i have a rank of 2 (implying $P \geq 3$), we can exchange the position of $\mathbf{w}_i$ and $\mathbf{w}_1$ such that $\mathbf{w}_1$ has a rank of 2. Then, there exist 2×2 matrices $\mathbf{A}^f, f = 2, \ldots, F$, such that Equation (4.21) holds. Then if $\mathbf{A}^f, f = 2, \ldots, F$, are all orthonormal matrices, the shape is not determined and the rotations can be uniquely determined as follows:

$$\mathbf{R}_f = \begin{bmatrix} \mathbf{A}_f & & 0 \\ & & 0 \\ 0 & 0 & |\mathbf{A}_f| \end{bmatrix}, \tag{4.47}$$

where $|\mathbf{A}^f|$ is the determinant of $\mathbf{A}^f$, which is either 1 or -1. Otherwise, the normalized 3D points lie on a plane whose equation at the first view can be written as (4.2). Then when the $\mathbf{C}$ matrix has a rank of 3 (implying $F \geq 4$), the plane for the first view can be determined to within a mirror uncertainty. When $\mathbf{C}$ matrix has a rank of 2 (implying $F \geq 3$), in general there are four plane solutions for the first view, and if and only if Equation (4.39) is satisfied for two independent equations of the form (4.31), the number of solutions can be reduced to two. If $\mathbf{C}$ has a rank of 1, then there are infinitely many solutions for the plane at the first view. For each plane solution for the first view, the rotation $\mathbf{R}_f$ between the fth and the first views has up to two solutions determined from Equation (4.27).†

A physical interpretation of the above conditions has not been obtained, though this shortcoming does not affect the usefulness of the conditions in the process of shape and motion estimation.

5 When rank(W) $\leq$ 2 and rank($\mathbf{w}_i$) = 1 for All i

Now let us consider the situation where each entry $\mathbf{w}_i$ of the measurement matrix $\mathbf{W}$ has a rank of 1. Obviously only when the image points in all views are colinear in the image plane, will such situation occur. This situation is quite similar to but more general than the one discussed by Tomasi and Kanade (Tomasi and Kanade 1990).

Clearly, if the image points are colinear in all views, the 3-D points must lie on a plane orthogonal to the image plane. This corresponds to the situation $n_3 = 0$ in the planar surface case that is just discussed above. Therefore the following equation holds for normalized image data for each view:

$$[n_1 \quad n_2]\Xi_i = 0. \tag{5.1}$$

When such a situation occurs, we can first solve for the rotations about the optical axis and then align the lines in different views all to the same direction, say, to the X axis by rotating the lines about the optical axis. Then the three dimensional motion problem is reduced to the two dimensional motion problem to be discussed below.

Consider the two dimensional case of the motion problem under orthographic projection. To make the notation consistent with the three dimensional case, we consider, without loss of generality, a special case of the three dimensional motion problem in which the rotations are all about the Y axis. Then the two-view motion equation (2.3) for normalized data reduces to

$$x'_i = [\cos\theta \quad -\sin\theta]\begin{bmatrix} x_i \\ Z_i \end{bmatrix}, \tag{5.2}$$

where θ is the rotation angle about the Y axis. The information on the other dimension (Y axis) does not contribute to the solution of rotation and shape, but only to the solution of the translation component in Y direction. Given P trajectories of points over F frames with the normalized coordinates on the X axis denoted by $x_i^f, i = 1, 2, \ldots, p, f = 1, 2, \ldots, F$, then we can obtain the following equation for long sequence motion

$$\begin{aligned} \Psi &= \begin{bmatrix} x_1^1 & \cdots & x_P^1 \\ & \cdots & \\ x_F^1 & \cdots & x_F^P \end{bmatrix}_{F\times P} \\ &= \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 \\ \cdots & \cdots \\ \cos\theta_F & -\sin\theta_F \end{bmatrix}_{F\times 2} [\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_P]_{2\times P} \\ &\equiv \mathbf{ms}, \end{aligned} \tag{5.3}$$

where $\theta_f, f = 1, 2, \ldots, F$, is the rotation angle about Y axis between the fth frame and the first frame, and $\mathbf{x}_p = [x_p, Z_p]^{\mathrm{T}}$ is the pth point in the X-Z plane. It follows that $\theta_1 = 0$. An $F \times 2$ matrix $\mathbf{m}$ and a $2 \times P$ matrix $\mathbf{s}$ are said to be a *motion factorization* of Ψ if and only if $\mathbf{m}$ can be represented in the following form

$$\mathbf{m} = \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 \\ \cdots & \cdots \\ \cos\theta_F & -\sin\theta_F \end{bmatrix} \tag{5.4}$$

with $\theta_1 = 0$. Again, the mirror uncertainty exists for this situation. That is, if $\mathbf{m}$ and $\mathbf{s}$ are a motion factorization of Ψ, so are $\mathbf{mJ}_2$ and $\mathbf{J}_2\mathbf{s}$, where $\mathbf{J}_2 = \mathrm{diag}(1, -1)$ is the two dimensional mirror reflection matrix. Now the problem is: given a measurement matrix Ψ, under what conditions can it have a unique mirror conjugate pair of motion factorization? The following theorem answers this question.

THEOREM 5.1. *For the two-dimensional case, if the normalized measurement matrix $\Psi_{F\times P}$ has a rank of 2, $\Psi_{F\times P}$ can be factorized into* $\mathbf{ly}$ *for some $F \times 2$ matrix* $\mathbf{l}$ *and some $2 \times P$ matrix* $\mathbf{y}$, *both of which have a rank of 2. Let*

$$\mathbf{l} = \begin{bmatrix} l_{11} & l_{12} \\ \cdots & \cdots \\ l_{F1} & l_{F2} \end{bmatrix}, \tag{5.5}$$

then if and only if

$$\mathbf{d} = \begin{bmatrix} l_{11}^2 & l_{11}l_{12} & l_{12}^2 \\ & \cdots & \\ l_{F1}^2 & l_{F1}l_{F2} & l_{F2}^2 \end{bmatrix} \tag{5.6}$$

has a rank of 3, or if and only if

$$\mathbf{d}' = \begin{bmatrix} \cos^2\theta_1 & -\cos\theta_1\sin\theta_1 & \sin^2\theta_1 \\ & \cdots & \\ \cos^2\theta_F & -\cos\theta_F\sin\theta_F & \sin^2\theta_F \end{bmatrix} \tag{5.7}$$

has a rank of 3, both the motion and shape are determined up to a mirror uncertainty, where θ_i is the true rotation angle for the ith view. Otherwise, both the motion and shape are not determined.

Proof. That Ψ can be factorized into $\mathbf{ly}$ for some $F\times 2$ matrix $\mathbf{l}$ and some $2\times P$ matrix $\mathbf{y}$ is obvious from the motion equation (5.3). Similar to the three dimensional case discussed in Theorem 3.1, to find a motion factorization, we need to find a 2×2 invertible matrix $\mathbf{a}$ such that $\mathbf{la}$ is a motion matrix of the form (5.4) with $\theta_1 = 0$. Let

$$\mathbf{p} = \mathbf{aa}^{\mathrm{T}} = \begin{bmatrix} p_1 & p_2 \\ p_2 & p_3 \end{bmatrix}, \tag{5.8}$$

To make $\mathbf{la}$ of the form (5.4) we then have the following equation for $\mathbf{p}$:

$$\mathbf{d}_{F\times 3} \cdot e(\mathbf{p}) = \mathbf{b}_{F\times 1}, \tag{5.9}$$

where $\mathbf{d}$ is represented in Equation (5.6) and

$$e(\mathbf{p}) = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{F\times 1}. \tag{5.10}$$

It is straightforward that if and only if **d** has a rank of 3, **p** can be uniquely determined from equation (5.9). For noiseless data, **d** and **d**$'$ have the same rank. Therefore, if and only if **d**$'$ has a rank of 3, **p** can be uniquely determined. After **p** is determined uniquely, matrix **a** can be determined up to an orthonormal matrix **u** from **p**. That is, if **a** is a symmetric factorization of **p** such that $\mathbf{aa}^{\mathrm{T}} = \mathbf{p}$, so is **au** for any 2×2 orthonormal matrix **u**. Assume one of the symmetric factorization **a** of **p** is obtained, e.g., using the SVD technique described in the Appendix. We then need to find the orthonormal matrix **u** such that **lau** is a motion matrix of the form (5.4) with $\theta_1 = 0$. Let

$$[\, l_{11} \quad l_{12}\,]\mathbf{a} = [\, x_1 \quad x_2\,], \qquad (5.11)$$

then **u** is to be determined from

$$[x_1 \quad x_2]\mathbf{u} = [1 \quad 0], \quad \text{and} \quad \mathbf{u}^{\mathrm{T}}\mathbf{u} = \mathbf{I}_2, \qquad (5.12)$$

where $\mathbf{I}_2$ is a 2×2 identity matrix . The solution of the above equation is given in closed form by the following formula

$$\mathbf{u} = \frac{1}{x_1^2 + x_2^2} \begin{bmatrix} x_1 & -x_2 \\ x_2 & x_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \pm 1 \end{bmatrix}, \qquad (5.13)$$

where the $\pm$ sign corresponds to the mirror uncertainty. Since for noiseless data, at least one solution must exist, in the above equation, we must have $x_1^2 + x_2^2 \neq 0$. Therefore, **u** can always be determined to within two sets as long as **p** can be determined uniquely.

We have thus proved the theorem. Q.E.D.

It is easy to verify that matrix **d** or **d**$'$ has a rank of 3 if and only if there are three distinct angles θ_i in **d**$'$, or if and only if there are three distinct views, consistent with Ullman's proof (Ullman 1979).

Now let us consider the situation where Ψ has a rank of 1. From equation (5.3), we know that there are only two ways in which Ψ can have a rank of 1: either the motion matrix **m** has a rank of 1, or the shape matrix **s** has a rank of 1. In the first case,

$$\mathbf{m} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{bmatrix}, \quad \Psi = \begin{bmatrix} x_1^1 & \cdots & x_P^1 \\ & \cdots & \\ x_1^1 & \cdots & x_P^1 \end{bmatrix}. \qquad (5.14)$$

That is, the image points never move in the image sequences and the rows of Ψ are the same. It is obvious that the shape is not determined. However one has to be cautious to conclude that the motions are all zero. If the points in the X-Z plane are not colinear, then it is always true that the rotations have to be zero. However, if the points in the X-Z plane are on a line, then there are two positions for the line which yield identical projections on the X axis. But the probability that in the image sequence the points move back and forth between the two positions is zero. In the second case, the motion is not degenerate, but the points lie on a line in the motion plane (X-Z plane) and consequently

$$\mathbf{s} = \begin{bmatrix} x_1^1 \, \alpha_2 x_1^1 \cdots \alpha_P x_1^1 \\ Z_1^1 \alpha_2 Z_1^1 \cdots \alpha_P Z_1^1 \end{bmatrix},$$
$$\Psi = \begin{bmatrix} x_1^1 \, \alpha_2 x_1^1 \cdots \alpha_P x_1^1 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ x_F^1 \alpha_2 x_F^1 \cdots \alpha_P x_F^1 \end{bmatrix}. \qquad (5.15)$$

That is, the columns of **s** are proportional to each other and the same is true for Ψ. It is clear that in this case one column of Ψ provides all information existing in Ψ, implying there are only F independent equations for $F + 1$ unknowns (F rotation angles and 1 depth). Therefore, the motion and shape (the orientation of the line) are both undetermined. In summary, we have the following theorem.

THEOREM 5.2. *When Ψ has a rank of* 1, *the shape (i.e. the depths of the points) is not determined. In this case, if the rows of Ψ are identical, then there are two possibilities: either the original* 3D *points are not colinear and the motion is zero, or the points are colinear and possibly move between two mirror symmetric positions; if the rows of Ψ are not identical but the columns of Ψ are proportional to each other, then the motion is not determined.*†

6 Summary

In this paper we have obtained new forms of necessary and sufficient conditions for determining the shape and motion to the minimum number of solutions from point trajectories under orthographic projection. All surface configurations of the points, including general, planar, and colinear cases, have been considered. Unlike Ullman's condition which is expressed in terms of the surface situation and the number of distinct views, the new conditions are expressed in terms of the image data and hence can be used in any practical algorithms to determine the uniqueness of the solution. These conditions enhance our understanding of the motion problem under orthographic projection. We have formally proved the mirror uncertainty which always exists for orthographic projection. All proofs are constructive so that they define a robust factorization algorithm for determining the uniqueness of the solution as well as the solution of shape and motion itself.

Appendix SVD Based Factorization

In this appendix we briefly discuss the singular value decomposition (SVD) factorization method that is needed to factorize some matrices. There are standard methods (Golub 1971) and commercial packages (IMSL 1987) for the SVD factorization.

Given a matrix $\mathbf{W}_{2F\times P}$ of rank $k, k = 1, 2, 3$, we now find a factorization $\mathbf{L}_{2F\times k}$ and $\mathbf{Y}_{k\times P}$ such that $\mathbf{W} = \mathbf{LY}$. Using the SVD technique, $\mathbf{W}$ can be factorized into the following form

$$\mathbf{W} = \mathbf{U}_{2F\times N}\Lambda_{N\times N}\mathbf{V}_{N\times P}, \tag{A.1}$$

where $N = \text{minimum}(2F, P)$, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)$ with singular values $\lambda_1 \geq \cdots \geq \lambda_N$, and $\mathbf{U}$ and $\mathbf{V}$ are two matrices of dimensions $2F \times N$ and $N \times P$ respectively. When $\mathbf{W}$ has a rank of k, then exactly the first k singular values are not zero. Therefore, let $\mathbf{L}_{2F\times k}$ be the first k columns of $\mathbf{U}$, and $\mathbf{R}_{k\times P}$ be the first k rows of $\mathbf{V}$, and $\Lambda_k = \text{diag}(\lambda_1, \ldots, \lambda_k)$, then it is clear that

$$\mathbf{W} = \mathbf{L}\Lambda_k\mathbf{R} = \mathbf{L}(\Lambda_k\mathbf{R}) = (\mathbf{L}\Lambda_k)\mathbf{R}. \tag{A.2}$$

$\mathbf{L}$ and $\Lambda_k\mathbf{R}$ or $\mathbf{L}\Lambda_k$ and $\mathbf{R}$ consist of a desired decomposition.

Now let us discuss how to factorize a symmetric matrix $\mathbf{P}_{N\times N}$ into the form $\mathbf{AA}^{\mathrm{T}}$. When $\mathbf{P}$ is symmetric, there exists an orthonormal matrix $\mathbf{U}$ such that

$$\mathbf{P} = \mathbf{U}\Lambda\mathbf{U}^{\mathrm{T}}, \tag{A.3}$$

where Λ is a diagonal matrix, or $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)$. Let $\Lambda^{\frac{1}{2}} = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_N})$. Then $\mathbf{A} = \mathbf{U}\Lambda^{\frac{1}{2}}$ is one of the desired factors such that $\mathbf{P} = \mathbf{AA}^{\mathrm{T}}$.

Acknowledgment

We would like to thank the anonymous reviewers for their very helpful comments.

References

J. Aloimonos and C. Brown, "Perception of structure from motion: I: Optical flow vs. discrete displacements, II: Lower bound results," *Proc. Int. Joint Conf. Pattern Recognition*, pp. 510–517, 1986.

C. Debrunner and N. Ahuja, "A direct data approximation based motion estimation algorithm," *Proc. 10th Int. Conf. Pattern Recognition*, pp. 384–389, 1990.

C. Debrunner and N. Ahuja, "Motion and structure factorization and segmentation of long multiple image sequences," *Proc. 2nd European Conf. Computer Vision*, pp. 217–221, 1992a.

C. Debrunner and N. Ahuja, "Motion and structure factorization and segmentation of long multiple image sequences," *Proc. Image Understanding Workshop*, pp. 543–547, 1992b.

G. Golub and C. Reinsch, *Singular Value Decomposition and Least Squares Solutions*, vol. 2. New York, NY: Springer Verlag, 1971.

X. Hu and N. Ahuja, "Motion estimation under orthographic projection," *IEEE Trans. Robot. and Automat.*, vol. 7, no. 6, pp. 848–853, 1991a.

X. Hu and N. Ahuja, "Sufficient conditions for double or unique solution of motion and structure," *Proc. Int. Conf. Acoust., Speech, and Signal Process.*, vol. 4, pp. 2445–2448, 1991b.

T. Huang and C. Lee, "Motion and structure from orthographic projections," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. PAMI-11, no. 5, pp. 536–540, 1989.

IMSL, *User's Manual: Fortran Subroutines for Mathematical Applications*. Houston, TX:IMSL, Inc., 1987.

K. Kanatani, "Structure and motion from optical flow under orthographic projection," *Comput. Vis., Graph., Image Process.*, no. 35, pp. 181–199, 1986.

C. Tomasi and T. Kanade, "Shape and motion without depth," *Proc. 3rd Int. Conf. Computer Vision*, pp. 91–95, 1990.

C. Tomasi and T. Kanade, "The factorization method for the recovery of shape and motion from image streams," *Proc. DARPA Image Understanding Workshop*, pp. 459–472, 1992.

S. Ullman, "The interpretation of visual motion," Ph.D. dissertation, MIT, 1977.

S. Ullman, *The Interpretation of Visual Motion*. Cambridge, Massachusetts, and London: The MIT Press, 1979.