Chapter 10

Projective Reconstruction of Shape and Motion Using Invariant Theory

Eduardo Bayro Corrochano and Vladimir Banarer

10.1 Introduction

In this chapter we present a geometric approach for the computation of shape and motion using projective invariants in the geometric algebra framework [6, 7].

In the last years researchers have developed diverse methods to compute projective invariants using n uncalibrated cameras [1, 2, 4, 8]. Different approaches for projective reconstruction have utilized the projective depth [13, 14], projective invariants [4] and factorization methods [11, 15, 16]. The factorization methods require the projective depth. The contribution of this paper is the application of projective invariants depending on the fundamental matrix or trifocal tensor to compute the projective depths. Using these projective depths we initialize the projective reconstruction procedure to compute shape and motion. We also illustrate the application of algebra of incidence for the development of geometric inference rules to complete the 3D data. The experimental part shows projective reconstruction of shape and motion using both simulated and real images.

The organization of the chapter is as follows: section two explains the generation and computation of projective invariants using two and three uncalibrated cameras. We test their performance using both simulated and real images. Section three presents the computation of the projective depth using projective invariants in terms of the trifocal tensor. The treatment of projective reconstruction and the role of the algebra of incidence to complete the 3–D shape is given in section four. The conclusion part follows.

10.2 3-D Projective Invariants from Multiple Views

This section presents the point and line projective invariants computable by means of n uncalibrated cameras. We begin with the generation of geometric invariants using the Plücker–Grassmann quadratic relation. We give a geometric interpretation of the cross–ratio in the 3–D space and in the image plane. We compute then projective invariants using two and three cameras.

10.2.1 Generation of geometric projective invariants

We choose for the visual projective space P^3 the geometric algebra $\mathcal{G}_{1,3,0}$ and for the image or projective plane P^2 the geometric algebra $\mathcal{G}_{3,0,0}$. Any 3D point is written in $\mathcal{G}_{1,3,0}$ as $\mathbf{X}_n = X_n\gamma_1 + Y_n\gamma_2 + Z_n\gamma_3 + W_n\gamma_4$ and its projected image point in $\mathcal{G}_{3,0,0}$ as $\boldsymbol{x}_n = x_n\sigma_1 + y_n\sigma_2 + z_n\sigma_3$, where $x_n = X_n/W_n$, $y_n = Y_n/W_n$, $z_n = Z_n/W_n$. The 3–D projective basis consists of four basis points and a fifth one for normalization: $\boldsymbol{X}_1 = [1,0,0,0]^T$, $\boldsymbol{X}_2 = [0,1,0,0]^T$, $\boldsymbol{X}_3 = [0,0,1,0]^T$, $\boldsymbol{X}_4 = [0,0,0,1]^T$, $\boldsymbol{X}_5 = [1,1,1,1]^T$ and the 2–D projective basis comprises three basis points and one for normalization: $\boldsymbol{x}_1 = [1,0,0]^T$, $\boldsymbol{x}_2 = [0,1,0]^T$, $\boldsymbol{x}_3 = [0,0,1]^T$, $\boldsymbol{x}_4 = [1,1,1]^T$. Using them we can express in terms of brackets the 3D projective coordinates X_n, Y_n, Z_n for any 3D point, as well as its 2D projected coordinates x_n, y_n

$$\frac{X_n}{W_n} = \frac{[234n][1235]}{[2345][123n]}, \quad \frac{Y_n}{W_n} = \frac{[134n][1235]}{[1345][123n]}, \quad \frac{Z_n}{W_n} = \frac{[124n][1235]}{[1245][123n]}. \tag{2.1}$$

$$\frac{x_n}{w_n} = \frac{[23n][124]}{[234][12n]}, \quad \frac{y_n}{w_n} = \frac{[13n][124]}{[134][12n]}. \tag{2.2}$$

These equations are projective invariants relations and they can be used for example, to compute the position of a moving camera.

The projective structure and its projection on the 2–D image is related according to the following geometric constraint

$$\begin{pmatrix} 0 & w_5Y_5 & -y_5Z_5 & (y_5-w_5)W_5 \\ w_5X_5 & 0 & -x_5Z_5 & (x_5-w_5)W_5 \\ 0 & w_6Y_6 & -y_6Z_6 & (x_5-w_5)W_5 \\ 0 & w_6Y_6 & -y_6Z_6 & (y_6-w_6)W_6 \\ w_6X_6 & 0 & -x_6Z_6 & (x_6-w_6)W_6 \\ 0 & w_7Y_7 & -y_7Z_7 & (y_7-w_7)W_7 \\ w_7X_7 & 0 & -x_7Z_7 & (x_7-w_7)W_7 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{pmatrix} \begin{pmatrix} X_0^{-1} \\ Y_0^{-1} \\ Z_0^{-1} \\ W_0^{-1} \end{pmatrix} = 0, \tag{2.3}$$

where X_0, Y_0, Z_0 W_0 are the coordinates of the view point. Since the matrix is of rank < 4, any determinant of four rows becomes a zero. Considering

$(X_5, Y_5, Z_5 W_5) = (1, 1, 1, 1)$ as a normalizing point and taking the determinant formed by the first four rows of equation (2.3) we get the geometric constraint equation involving six points pointed out by Quan [12]

$$\begin{aligned}&(w_5y_6 - x_5y_6)X_6Z_6 + (x_5y_6 - x_5w6)X_6W_6 + (x_5w_6 - y_5w6)X_6Y_6 + \\ &+(y_5x_6 - w_5x6)Y_6Z_6 + (y_5w_6 - y_5x6)Y_6W_6 + \\ &\quad +(w_5x_6 - w_5y_6)Z_6W_6 = 0\end{aligned} \tag{2.4}$$

Carlsson [3] showed that the equation (2.4) can be also derived using the *Plücker–Grassmann relations*. This can be computed as the *Laplace expansion* of the 4×8 rectangular matrix involving the same six points as above

$$\begin{aligned}&[\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_4, \boldsymbol{X}_5, \boldsymbol{X}_5, \boldsymbol{X}_6, \boldsymbol{X}_7] = [\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3] \\ &[\boldsymbol{X}_4, \boldsymbol{X}_5, \boldsymbol{X}_6, \boldsymbol{X}_7] - [\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_4][\boldsymbol{X}_3, \boldsymbol{X}_5, \boldsymbol{X}_6, \boldsymbol{X}_7] + \\ &+[\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_5][\boldsymbol{X}_3, \boldsymbol{X}_4, \boldsymbol{X}_6, \boldsymbol{X}_7] - [\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_6] \\ &[\boldsymbol{X}_3, \boldsymbol{X}_4, \boldsymbol{X}_5, \boldsymbol{X}_7] + [\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_7][\boldsymbol{X}_3, \boldsymbol{X}_4, \boldsymbol{X}_5, \boldsymbol{X}_6] = 0.\end{aligned} \tag{2.5}$$

Using four functions like equation (2.5) in terms of the permutations of six points as indicated by their sub–indices in the table below

$\boldsymbol{X}_0$	$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	$\boldsymbol{X}_3$	$\boldsymbol{X}_4$	$\boldsymbol{X}_5$	$\boldsymbol{X}_6$	$\boldsymbol{X}_7$
0	1	5	1	2	3	4	5
0	2	6	1	2	3	4	6
0	3	5	1	2	3	4	5
0	4	6	1	2	3	4	6

we get an expression where the brackets that have two identical points vanish

$$\begin{aligned}&[0152][1345] - [0153][1245] + [0154][1235] = 0, \\ &[0216][2346] - [0236][1246] + [0246][1236] = 0, \\ &[0315][2345] + [0325][1345] + [0345][1235] = 0, \\ &[0416][2346] + [0426][1346] - [0436][1246] = 0.\end{aligned} \tag{2.6}$$

It is easy to show that the brackets of image points can be written in the form $[\boldsymbol{x}_i\boldsymbol{x}_j\boldsymbol{x}_k] = w_iw_jw_k[K][\boldsymbol{X}_0\boldsymbol{X}_i\boldsymbol{X}_j\boldsymbol{X}_k]$, where [K] is the matrix of the intrinsic parameters [10]. Now if we express in equations (2.6) all the brackets which have the point $\boldsymbol{X}_0$ in terms of the brackets of image points and organize all the bracket products as a 4×4 matrix we get the singular matrix

$$\begin{pmatrix} 0 & [125][1345] & [135][1245] & [145][1235] \\ [216][2346] & 0 & [236][1246] & [246][1236] \\ [315][2345] & [325][1345] & 0 & [345][1235] \\ [416][2346] & [426][1346] & [436][1246] & 0. \end{pmatrix} \tag{2.7}$$

Here the scalars $w_i w_j w_k[K]$ of each matrix entry cancel each other. Now after taking the determinat of this matrix and rearrange the terms conveniently, we obtain the following useful bracket polynomial

$$\begin{aligned}
&[125][346]\Big[1236\Big]\Big[1246\Big]\Big[1345\Big]\Big[2345\Big] - \\
&[126][345]\Big[1235\Big]\Big[1245\Big]\Big[1346\Big]\Big[2346\Big] + \\
&[135][246]\Big[1236\Big]\Big[1245\Big]\Big[1346\Big]\Big[2345\Big] - \\
&[136][245]\Big[1235\Big]\Big[1246\Big]\Big[1345\Big]\Big[2346\Big] + \\
&[145][236]\Big[1235\Big]\Big[1246\Big]\Big[1346\Big]\Big[2345\Big] - \\
&[146][235]\Big[1236\Big]\Big[1245\Big]\Big[1345\Big]\Big[2346\Big] = 0,
\end{aligned} \tag{2.8}$$

Surprisingly this bracket expression is exactly the *shape constraint* for six points given by Quan [12]

$$i_1 I_1 + i_2 I_2 + i_3 I_3 + i_4 I_4 + i_5 I_5 + i_6 I_6 = 0, \tag{2.9}$$

where $i_1 = [125][346]$, $i_2 = [126][345]$, ..., $i_6 = [146][235]$ and
$I_1 = [1236][1246][1345][2345]$, $I_2 = [1235][1245][1346][2346]$, ...,
$I_6 = [1236][1245][1345][2346]$ are the the relative linear invariants in P^2 and P^3 respectively. Using the shape constraint we are now ready to generate invariants for different purpose.

Let us illustrate this with an example. As shown in the Figure 10.1 there is a configuration of six points which indicates whether or not the end–effector is grasping properly.

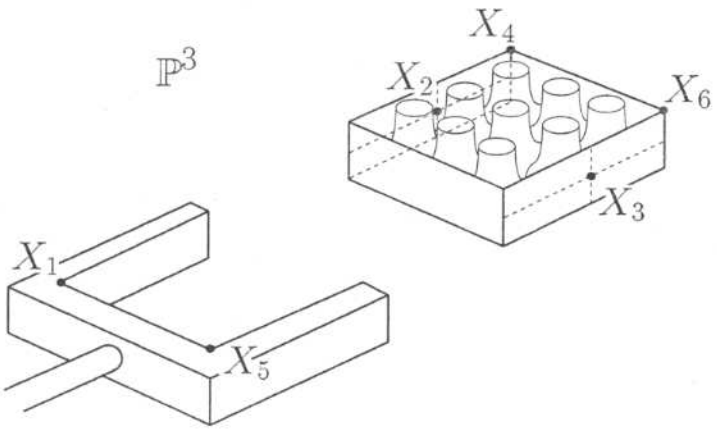


FIGURE 10.1. Grasping a box.

To test this situation we can use an invariant generated from the constraint of equation (2.8). In this particular situation we recognize two planes:

[1235]=0 and [2346]=0. Substituting these six points in equation (2.8) we make some brackets vanish reducing the equation to

$$\begin{aligned}&[125][346]\Big[1236\Big]\Big[1246\Big]\Big[1345\Big]\Big[2345\Big] - \\ &-[135][246]\Big[1236\Big]\Big[1245\Big]\Big[1346\Big]\Big[2345\Big] = 0\end{aligned} \tag{2.10}$$

$$[125][346]\Big[1246\Big]\Big[1345\Big] - [135][246]\Big[1245\Big]\Big[1346\Big] = 0 \tag{2.11}$$

or

$$\begin{aligned} Inv &= \frac{(\mathbf{X}_1 \wedge \mathbf{X}_2 \wedge \mathbf{X}_4 \wedge \mathbf{X}_5) I_4^{-1} (\mathbf{X}_1 \wedge \mathbf{X}_3 \wedge \mathbf{X}_4 \wedge \mathbf{X}_6) I_4^{-1}}{(\mathbf{X}_1 \wedge \mathbf{X}_2 \wedge \mathbf{X}_4 \wedge \mathbf{X}_6) I_4^{-1} (\mathbf{X}_1 \wedge \mathbf{X}_3 \wedge \mathbf{X}_4 \wedge \mathbf{X}_5) I_4^{-1}} \\ &= \frac{(\boldsymbol{x}_1 \wedge \boldsymbol{x}_2 \wedge \boldsymbol{x}_5) I_3^{-1} (\boldsymbol{x}_3 \wedge \boldsymbol{x}_4 \wedge \boldsymbol{x}_6) I_3^{-1}}{(\boldsymbol{x}_1 \wedge \boldsymbol{x}_3 \wedge \boldsymbol{x}_5) I_3^{-1} (\boldsymbol{x}_2 \wedge \boldsymbol{x}_4 \wedge \boldsymbol{x}_6) I_3^{-1}}. \end{aligned} \tag{2.12}$$

In this equation any bracket of P^3 after the projective mapping fulfills

$$\begin{aligned}&(\mathbf{X}_1 \wedge \mathbf{X}_2 \wedge \mathbf{X}_4 \wedge \mathbf{X}_5) I_4^{-1} \equiv \\ &\quad W_1 W_2 W_4 W_5 \{(\boldsymbol{x}_2 - \boldsymbol{x}_1) \wedge (\boldsymbol{x}_4 - \boldsymbol{x}_1) \wedge (\boldsymbol{x}_5 - \boldsymbol{x}_1)\} I_3^{-1},\end{aligned} \tag{2.13}$$

The constraint (2.8) makes always sure that the $W_i W_j W_k W_l$ constants are canceled. Furthermore, we can interpret the invariant Inv, the equivalent of the , in P^3 as ratios of volumes and in P^2 as rations of triangle areas

$$Inv = \frac{V_{1245} V_{1346}}{V_{1246} V_{1345}} = \frac{A_{125} A_{346}}{A_{135} A_{246}}. \tag{2.14}$$

In other words, we can also see this invariant in P^3 as the relation of 4-vectors or volumes built by points lying on a quadric which projected in P^2 represents an invariant build by areas of triangles encircled by conics.

For example utilizing this invariant we can check whether or not the grasper is holding the box correctly. Note that using the observed 3–D points in the image we can compute this invariant and see if the relation of the triangle areas corresponds with the appropriate relation for firm grasping, i.e. if the grasper is away the invariant has a different value from the required value when the points $\boldsymbol{X}_1$, $\boldsymbol{X}_5$ of the grasper are near to the objects points $\boldsymbol{X}_2$, $\boldsymbol{X}_3$.

10.2.2 Projective invariants using two views

Let us consider a 3–D *projective invariant* derived from the equation (2.8)

$$\boldsymbol{Inv_3} = \frac{[\mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_3 \mathbf{X}_4][\mathbf{X}_4 \mathbf{X}_5 \mathbf{X}_2 \mathbf{X}_6]}{[\mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_4 \mathbf{X}_5][\mathbf{X}_3 \mathbf{X}_4 \mathbf{X}_2 \mathbf{X}_6]}. \tag{2.15}$$

The computation of the bracket

$$[1234] = (\mathbf{X}_1 \wedge \mathbf{X}_2 \wedge \mathbf{X}_3 \wedge \mathbf{X}_4)\boldsymbol{I}_4^{-1} = ((\mathbf{X}_1 \wedge \mathbf{X}_2) \wedge (\mathbf{X}_3 \wedge \mathbf{X}_4))\boldsymbol{I}_4^{-1}$$

of four points from R^4, mapped to the cameras with the optical centers $\mathbf{A}_0$ and $\mathbf{B}_0$, suggests to use the binocular model based on incidence algebra as introduced in chapter 7. Defining the lines

$$\begin{aligned} \boldsymbol{L}_{12} &= \mathbf{X}_1 \wedge \mathbf{X}_2 = (\mathbf{A}_0 \wedge \boldsymbol{L}_{12}^A) \vee (\mathbf{B}_0 \wedge \boldsymbol{L}_{12}^B) \\ \boldsymbol{L}_{34} &= \mathbf{X}_3 \wedge \mathbf{X}_4 = (\mathbf{A}_0 \wedge \boldsymbol{L}_{34}^A) \vee (\mathbf{B}_0 \wedge \boldsymbol{L}_{34}^B) \end{aligned}$$

where lines $\boldsymbol{L}_{ij}^A$ and $\boldsymbol{L}_{ij}^B$ are mappings of the line $\boldsymbol{L}_{ij}$ to the two image planes, results in the following expression for the bracket

$$[1234] = [\mathbf{A}_0 \mathbf{B}_0 \mathbf{A}'_{1234} \mathbf{B}'_{1234}]. \tag{2.16}$$

Here $\mathbf{A}'_{1234}$ and $\mathbf{B}'_{1234}$ are the points of intersection of the lines $\boldsymbol{L}_{12}^A$ and $\boldsymbol{L}_{34}^A$ or $\boldsymbol{L}_{12}^B$ and $\boldsymbol{L}_{34}^B$, respectively. These points, lying in the image planes, can be expanded using the mappings of three points $\mathbf{X}_i$, say $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$, to the image planes, i.e. $\mathbf{A}_j$ and $\mathbf{B}_j$, $j = 1, 2, 3$, as projective basis, as follows

$$\begin{aligned} \mathbf{A}'_{1234} &= \alpha_{1234,1}\mathbf{A}_1 + \alpha_{1234,2}\mathbf{A}_2 + \alpha_{1234,3}\mathbf{A}_3 \\ \mathbf{B}'_{1234} &= \beta_{1234,1}\mathbf{B}_1 + \beta_{1234,2}\mathbf{B}_2 + \beta_{1234,3}\mathbf{B}_3. \end{aligned}$$

Then equation (15.73) from chapter 15 follows

$$[1234] = \sum_{i,j=1}^{3} \tilde{F}_{ij}\alpha_{1234,i}\beta_{1234,j} = \boldsymbol{\alpha}_{1234}^T \tilde{F} \boldsymbol{\beta}_{1234}, \tag{2.17}$$

where $\tilde{F}$ is the fundamental matrix given in terms of the projective basis, embedded in R^4 and $\boldsymbol{\alpha}_{1234} = (\alpha_{1234,1}, \alpha_{1234,2}, \alpha_{1234,3})$ and $\boldsymbol{\beta}_{1234} = (\beta_{1234,1}, \beta_{1234,2}, \beta_{1234,3})$ are corresponding points.

The ratio

$$\boldsymbol{Inv_{3F}} = \frac{(\boldsymbol{\alpha}^T{}_{1234}\tilde{F}\boldsymbol{\beta}_{1234})(\boldsymbol{\alpha}^T{}_{4526}\tilde{F}\boldsymbol{\beta}_{4526})}{(\boldsymbol{\alpha}^T{}_{1245}\tilde{F}\boldsymbol{\beta}_{1245})(\boldsymbol{\alpha}^T{}_{3426}\tilde{F}\boldsymbol{\beta}_{3426})} \tag{2.18}$$

is therefore seen to be an invariant using two cameras [2]. Note that equation (2.18) is invariant whatever values of the γ_4 components of the vectors $\mathbf{A}_i, \mathbf{B}_i, \mathbf{X}_i$ etc. are chosen. A confusion arises if we attempt to express the invariant of equation (2.18) in terms of what we actually observe, i.e. the homogeneous Cartesian image coordinates $\boldsymbol{a}'_i s$, $\boldsymbol{b}'_i s$ and the fundamental

matrix F calculated from these image coordinates. In order to avoid that it is necessary to transfer the computations of equation (2.18) carried out in R^4 to R^3. Let us explain now this procedure.

If we define $\tilde{F}$ by

$$\tilde{F}_{kl} = (\mathbf{A}_k \cdot \gamma_4)(\mathbf{B}_l \cdot \gamma_4) F_{kl} \tag{2.19}$$

then using the relationships $\boldsymbol{\alpha}_{ij} = \frac{\mathbf{A}'_i \cdot \gamma_4}{\mathbf{A}_j \cdot \gamma_4} \boldsymbol{a}_{ij}$ and $\boldsymbol{\beta}_{ij} = \frac{\mathbf{B}'_i \cdot \gamma_4}{\mathbf{B}_j \cdot \gamma_4} \boldsymbol{b}_{ij}$ it follows that

$$\boldsymbol{\alpha}_{ik} \tilde{F}_{kl} \boldsymbol{\beta}_{il} = (\mathbf{A}'_i \cdot \gamma_4)(\mathbf{B}'_i \cdot \gamma_4) \boldsymbol{a}_{ik} F_{kl} \boldsymbol{b}_{il}. \tag{2.20}$$

If F is estimated by some method, then an $\tilde{F}$ defined as in equation (2.19) will also act as a *fundamental matrix* or *bilinear constraint* in R^4. Now let us look again at the invariant $\boldsymbol{Inv_{3F}}$. According to the above considerations, we can write the invariant as

$$\boldsymbol{Inv_{3F}} = \frac{(\boldsymbol{a}^T{}_{1234} F \boldsymbol{b}_{1234})(\boldsymbol{a}^T{}_{4526} F \boldsymbol{b}_{4526}) \phi_{1234} \phi_{4526}}{(\boldsymbol{a}^T{}_{1245} F \boldsymbol{b}_{1245})(\boldsymbol{a}^T{}_{3426} F \boldsymbol{b}_{3426}) \phi_{1245} \phi_{3426}} \tag{2.21}$$

where $\phi_{pqrs} = (\mathbf{A}'_{pqrs} \cdot \gamma_4)(\mathbf{B}'_{pqrs} \cdot \gamma_4)$. Therefore we can see that the ratio of the terms $\boldsymbol{a}^T F \boldsymbol{b}$ which resembles the expression for the invariant in R^4 but uses only the observed coordinates and the estimated fundamental matrix will not be an invariant. Instead, we need to include the factors ϕ_{1234} etc., which do not cancel. It is relatively easy to show [1] that these factors can be formed as follows. Since $\boldsymbol{a}'_3$, $\boldsymbol{a}'_4$ and $\boldsymbol{a}'_{1234}$ are collinear, we can write $\boldsymbol{a}'_{1234} = \mu_{1234} \boldsymbol{a}'_4 + (1 - \mu_{1234}) \boldsymbol{a}'_3$. Then, by expressing $\mathbf{A}'_{1234}$ as the intersection of the line joining $\mathbf{A}'_1$ and $\mathbf{A}'_2$ with the plane through $\mathbf{A}_0, \mathbf{A}'_3, \mathbf{A}'_4$ we can use the projective split and equate terms so that they give

$$\frac{(\mathbf{A}'_{1234} \cdot \gamma_4)(\mathbf{A}'_{4526} \cdot \gamma_4)}{(\mathbf{A}'_{3426} \cdot \gamma_4)(\mathbf{A}'_{1245} \cdot \gamma_4)} = \frac{\mu_{1245}(\mu_{3426} - 1)}{\mu_{4526}(\mu_{1234} - 1)}. \tag{2.22}$$

Note that the values of μ are readily obtainable from the images. The factors $\mathbf{B}'_{pqrs} \cdot \gamma_4$ are found in a similar way so that if $\boldsymbol{b}'_{1234} = \lambda_{1234} \boldsymbol{b}'_4 + (1 - \lambda_{1234}) \boldsymbol{b}'_3$ etc., the overall expression for the invariant becomes

$$\begin{aligned} \boldsymbol{Inv_{3F}} \quad = \quad & \frac{(\boldsymbol{a}^T_{1234} F \boldsymbol{b}_{1234})(\boldsymbol{a}^T_{4526} F \boldsymbol{b}_{4526})}{(\boldsymbol{a}^T_{1245} F \boldsymbol{b}_{1245})(\boldsymbol{a}^T_{3426} F \boldsymbol{b}_{3426})} \cdot \\ & \frac{\mu_{1245}(\mu_{3426} - 1)}{\mu_{4526}(\mu_{1234} - 1)} \frac{\lambda_{1245}(\lambda_{3426} - 1)}{\lambda_{4526}(\lambda_{1234} - 1)}. \end{aligned} \tag{2.23}$$

As conclusion, given the coordinates of a set of 6 corresponding points in two image planes, where these 6 points are projections of arbitrary world points in general position, we can form 3–D projective invariants provided we have some estimate of F.

10.2.3 Projective invariant of points using three views

The technique used to form the 3–D projective invariants for two views can be straightforwardly extended to give expressions for invariants of three views. Considering four world points, $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$, or two lines $\mathbf{X}_1 \wedge \mathbf{X}_2$ and $\mathbf{X}_3 \wedge \mathbf{X}_4$, projected into three camera planes, we can write

$$\begin{aligned} \mathbf{X}_1 \wedge \mathbf{X}_2 &= (\mathbf{A}_0 \wedge \boldsymbol{L}_{12}^A) \vee (\mathbf{B}_0 \wedge \boldsymbol{L}_{12}^B) \\ \mathbf{X}_3 \wedge \mathbf{X}_4 &= (\mathbf{A}_0 \wedge \boldsymbol{L}_{34}^A) \vee (\mathbf{C}_0 \wedge \boldsymbol{L}_{34}^C). \end{aligned}$$

Once again, we can combine the above expressions so that they give to give an equation for the 4-vector $\mathbf{X}_1 \wedge \mathbf{X}_2 \wedge \mathbf{X}_3 \wedge \mathbf{X}_4$,

$$\begin{aligned} \mathbf{X}_1 \wedge \mathbf{X}_2 \wedge \mathbf{X}_3 \wedge \mathbf{X}_4 &= ((\mathbf{A}_0 \wedge \boldsymbol{L}_{12}^A) \vee (\mathbf{B}_0 \wedge \boldsymbol{L}_{12}^B)) \wedge ((\mathbf{A}_0 \wedge \boldsymbol{L}_{34}^A) \vee (\mathbf{C}_0 \wedge \boldsymbol{L}_{34}^C)) \\ &= (\mathbf{A}_0 \wedge \mathbf{A}_{1234}) \wedge ((\mathbf{B}_0 \wedge \boldsymbol{L}_{12}^B) \vee (\mathbf{C}_0 \wedge \boldsymbol{L}_{34}^C)). \end{aligned} \tag{2.24}$$

Writing the lines $\boldsymbol{L}_{12}^B$ and $\boldsymbol{L}_{34}^C$ in terms of the line coordinates we have $\boldsymbol{L}_{12}^B = \sum_{j=1}^{3} l_{12,j}^B \boldsymbol{L}_j^B$ and $\boldsymbol{L}_{34}^C = \sum_{j=1}^{3} l_{34,j}^C \boldsymbol{L}_j^C$.

It has been shown in chapter 15 that the components of the *trifocal tensor* (which plays the role of the fundamental matrix for 3 views), can be written in geometric algebra as

$$\tilde{T}_{ijk} = [(\mathbf{A}_0 \wedge \mathbf{A}_i) \wedge ((\mathbf{B}_0 \wedge \boldsymbol{L}_j^B) \vee (\mathbf{C}_0 \wedge \boldsymbol{L}_k^C))] \tag{2.25}$$

so that from equation (2.24) it can be derived:

$$[\mathbf{X}_1 \wedge \mathbf{X}_2 \wedge \mathbf{X}_3 \wedge \mathbf{X}_4] = \sum_{i,j,k=1}^{3} \tilde{T}_{ijk} \alpha_{1234,i} l_{12,j}^B l_{34,k}^C = \tilde{T}(\boldsymbol{\alpha}_{1234}, \boldsymbol{L}_{12}^B, \boldsymbol{L}_{34}^C) \tag{2.26}$$

The invariant $\boldsymbol{Inv_3}$ can then be expressed as

$$\boldsymbol{Inv_{3T}} = \frac{\tilde{T}(\boldsymbol{\alpha}_{1234}, \boldsymbol{L}_{12}^B, \boldsymbol{L}_{34}^C) \tilde{T}(\boldsymbol{\alpha}_{4526}, \boldsymbol{L}_{25}^B, \boldsymbol{L}_{26}^C)}{\tilde{T}(\boldsymbol{\alpha}_{1245}, \boldsymbol{L}_{12}^B, \boldsymbol{L}_{45}^C) \tilde{T}(\boldsymbol{\alpha}_{3426}, \boldsymbol{L}_{34}^B, \boldsymbol{L}_{26}^C)}. \tag{2.27}$$

Note that the factorization must be done so that the same line factorizations occur in both the numerator and denominator. Therefore we have an expression for invariants in three views that is a direct extension of the invariants for two views. Forming the above invariant from observed quantities we note, as before, that some correction factors will be necessary – equation (2.27) is given above in terms of R^4 quantities. Fortunately, this is quite straightforward. Regarding the results of previous section, we can simply consider the $\boldsymbol{\alpha}'s$ terms in equation (2.27) as not observable quantities, conversely the line terms like $\boldsymbol{L}_{12}^B, \boldsymbol{L}_{34}^C$ are indeed observed quantities.

As a result, the expression has to be modified using partially the coefficients computed in previous section and for the unique four combinations of three cameras their invariant equations read

$$\boldsymbol{Inv_{3T}} = \frac{T(\boldsymbol{a}_{1234}, \boldsymbol{l}^B_{12}, \boldsymbol{l}^C_{34}) T(\boldsymbol{a}_{4526}, \boldsymbol{l}^B_{25}, \boldsymbol{l}^C_{26})}{T(\boldsymbol{a}_{1245}, \boldsymbol{l}^B_{12}, \boldsymbol{l}^C_{45}) T(\boldsymbol{a}_{3426}, \boldsymbol{l}^B_{34}, \boldsymbol{l}^C_{26})} \frac{\mu_{1245}(\mu_{3426} - 1)}{\mu_{4526}(\mu_{1234} - 1)}. \tag{2.28}$$

10.2.4 Comparison of the projective invariants

Invariants using F

0.000	0.590	0.670	0.460
	0	0.515	0.68
		0.59	0
			0.69

0.063	0.650	0.750	0.643
	0.67	0.78	0.687
		0.86	0.145
			0.531

0.148	0.600	0.920	0.724
	0.60	0.96	0.755
		0.71	0.97
			0.596

0.900	0.838	0.690	0.960
	0.276	0.693	0.527
		0.98	0.59
			0.663

Invariants using T

0.000	0.590	0.310	0.630
	0	0.63	0.338
		0.134	0.67
			0.29

0.044	0.590	0.326	0.640
	0	0.63	0.376
		0.192	0.67
			0.389

0.031	0.100	0.352	0.660
	0.031	0.337	0.67
		0.31	0.67
			0.518

0.000	0.640	0.452	0.700
	0.063	0.77	0.545
		0.321	0.63
			0.643

FIGURE 10.2. The distance matrices show the performance of the invariants by increasing Gaussian noise σ: 0.005, 0.015, 0.025 and 0.04.

This section shows simulations with synthetic data and computations using real images. The simulation was implemented in Maple.

The computation of the bilinearity matrix F and the trilinearity focal tensor T was done using a linear method. We believe that for the test purposes these are good enough. Four different sets of six points $S_i = \{\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \boldsymbol{X}_{i3}, \boldsymbol{X}_{i4}, \boldsymbol{X}_{i5}, \boldsymbol{X}_{i6}\}$, where $i = 1, .., 4$, were considered in the simulation and the only three possible invariants were computed for each

set $\{I_{1,i},I_{2,i},I_{3,i}\}$. Then, the invariants of each set were represented as 3–D vectors ($\mathbf{v}_i = [I_{1,i}, I_{2,i}, I_{3,i}]^T$). We computed four of these vectors that corresponded to four different sets of six points using two images for the F case and three images for the T case (first group of images); and for four of these vectors corresponding to the same point sets we used another two images for the F case or another three images for the T case (second group of images). The comparison of the invariants was done using Euclidean distances of the vectors $d(\mathbf{v}_i, \mathbf{v}_j) = (1 - |\frac{\mathbf{v}_i \cdot \mathbf{v}_j}{||\mathbf{v}_i|| ||\mathbf{v}_j||}|)^{\frac{1}{2}}$; this method was used for the same reason by [5].

Since in $d(\mathbf{v}_i, \mathbf{v}_j)$ we normalize the vectors $\mathbf{v}_i$ and $\mathbf{v}_j$, the distance $d(\mathbf{v}_i, \mathbf{v}_j)$ for any of them does lies between 0 and 1 and it does not vary when $\mathbf{v}_i$ or $\mathbf{v}_j$ is multiplied by a nonzero constant. The figure 10.2 shows a comparison table where each (i, j)-th entry represents the distance computed using $d(\mathbf{v}_i, \mathbf{v}_j)$ between the invariants of set S_i of the points extracted of the first group of images and the set S_j of the points yet using the second group of images. In the ideal case, the diagonal of the distance matrices should be zero, that means that the values of the computed invariants remain constant regardless of which group of images they were used for. The entries off the diagonal mean that we are comparing vectors composed of different coordinates ($\mathbf{v}_i = [I_{1,i}, I_{2,i}, I_{3,i}]^T$), thus they are not parallel and should be bigger than zero and if they are very different the value of $d(\mathbf{v}_i, \mathbf{v}_j)$ should be approximately 1. Now looking at the figure 10.2, we can clearly see that the performance of the invariants based on trilinearities is much better than that of those based on bilinearities, the diagonal values in the T case are in general closer to zero than in the F case and its entries off the diagonal are in general bigger values than in the F case.

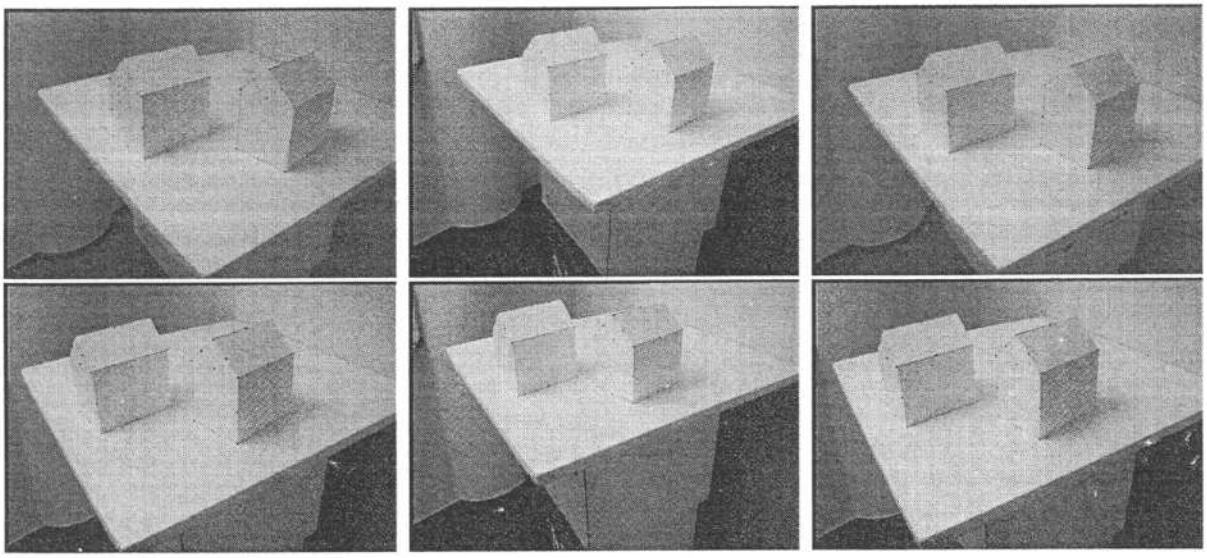


FIGURE 10.3. Image sequence taken during navigation by the binocular head of a mobile robot. The upper row shows the left camera images and the lower one shows the right camera ones.

In the case of real images we use a sequence of images taken by a moving robot equipped with a binocular head. The figure 10.3 shows three images of the left eye in the upper row and below these of the right eye respectively. We took image couples, one from the left and one from the right for the invariants using F and two of one eye and one of the other for the invariant using T. From the image we took 38 points semi–automatically and we selected now six sets of points. In each set the points are in general position. Three invariants of each set were computed and the comparison tables were obtained similarly to the previous experiment, see figure 10.4.

using F

0.04	0.79	0.646	0.130	0.679	0.89
	0.023	0.2535	0.278	0.268	0.89
		0.0167	0.723	0.606	0.862
			0.039	0.808	0.91
				0.039	0.808
					0.039

using T

0.021	0.779	0.346	0.930	0.759	0.81
	0.016	0.305	0.378	0.780	0.823
		0.003	0.83	0.678	0.97
			0.02	0.908	0.811
				0.008	0.791
					0.01

FIGURE 10.4. The distance matrices show the performance of the computed invariants using bilinearities (top) and trilinearities (bottom) for the image sequence.

This shows again that the approach to compute the invariants using trilinearities is much more robust than the one using bilinearities, as expected from the theoretical point of view.

10.3 Projective Depth

In a geometric sense the *projective depth* can be seen as the relation between the distance regarding the view center of a 3–D point $\boldsymbol{X}_i$ and the focal distance f as depicted in figure 10.5.

Let us derive the projective depth from a projective mapping. According to the pinhole model explained in chapter 15 the coordinates of a point in

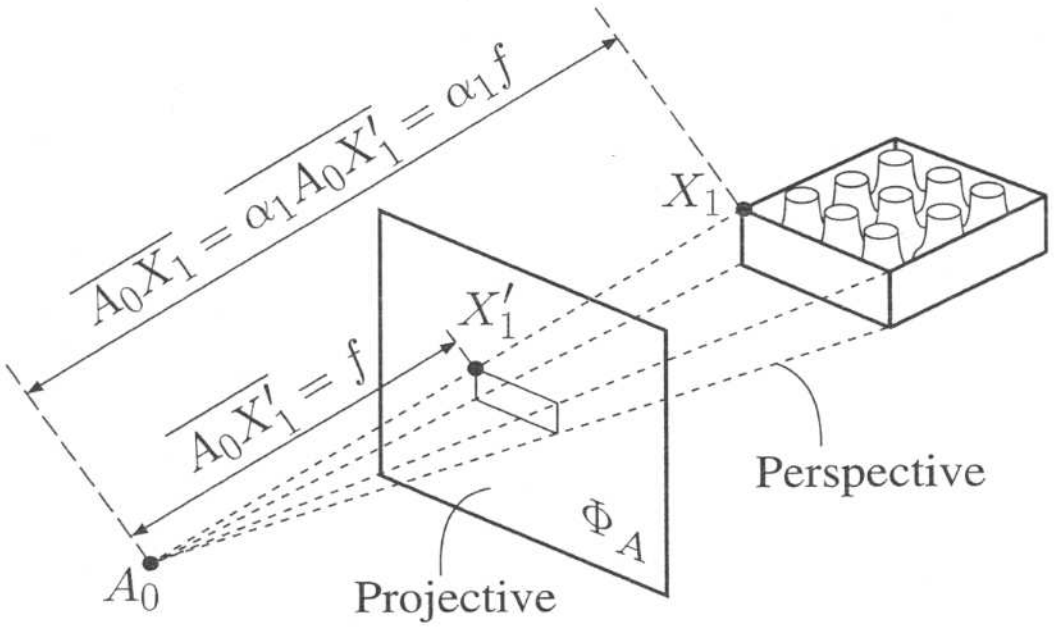


FIGURE 10.5. Geometric interpretation of the projective depth.

the image plane is the result of the projection of the 3–D point to the three optical planes ϕ_A^1, ϕ_A^2, ϕ_A^3. They are spanned by a trivector basis $\gamma_i, \gamma_j, \gamma_k$ and the coefficients t_{ij}. This projective mapping in a matrix representation reads

$$\begin{aligned}\lambda \boldsymbol{x} &= \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \phi_A^1 \\ \phi_A^2 \\ \phi_A^3 \end{bmatrix} \boldsymbol{X} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \end{aligned} \tag{3.29}$$

where the projective scale factor is called λ. Note that the projective mapping is further expressed in terms of a f, rotation and translation components. Let us attach the world coordinates to the view center of the camera. The resultant projective mapping becomes

$$\lambda \boldsymbol{x} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \equiv P\boldsymbol{X}. \tag{3.30}$$

We can then compute straightforwardly

$$\lambda = Z. \tag{3.31}$$

The way how we compute the projective depth ($\equiv \lambda$) of a 3–D point appears simple using invariant theory, namely using equations (2.1). For that we select a basis system taking four 3–D points in general position $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{X}_3$, $\boldsymbol{X}_5$, the optical center of camera at the new position as the four point $\boldsymbol{X}_4$, and $\boldsymbol{X}_6$ as the 3–D point to be reconstructed. This has been depicted in figure 10.6.

Since we use the mapped points, we consider the *epipole* (mapping of the current view center) as the four point and the mapped sixth point as the point with the unknown depth. The other mapped basis points remain constant during the procedure.

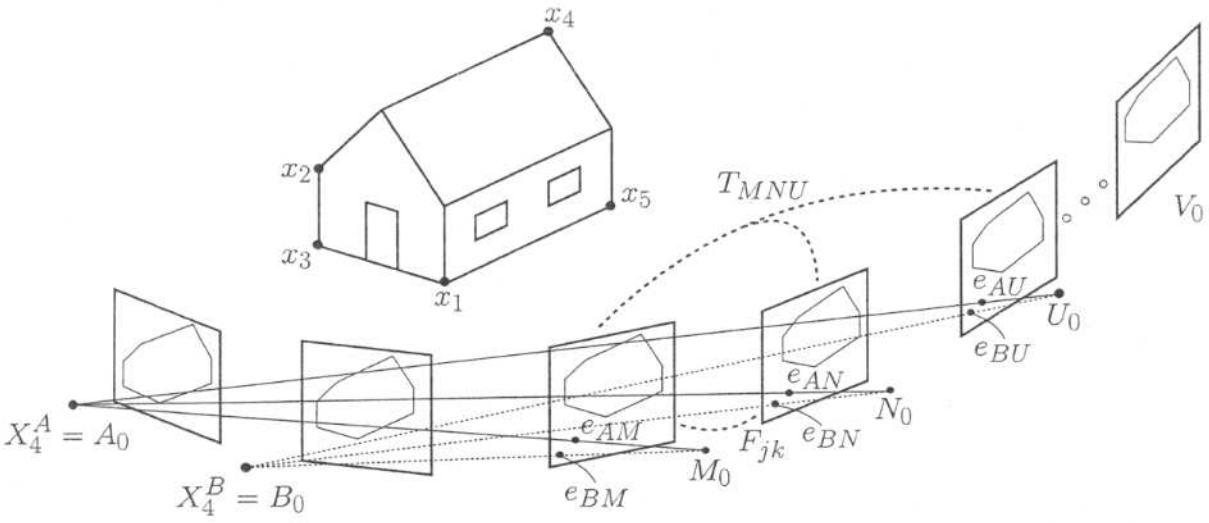


FIGURE 10.6. Computing the projective depths of n cameras.

According to equation (2.1), the tensor based expression for computing the third coordinate or projective depth of a point $\boldsymbol{X}_j$ ($= \boldsymbol{X}_6$) reads

$$\lambda_j = \frac{Z_j}{W_j} = \frac{T(\boldsymbol{a}_{124j}, \boldsymbol{l}^B_{12}, \boldsymbol{l}^C_{4j}) T(\boldsymbol{a}_{1235}, \boldsymbol{l}^B_{12}, \boldsymbol{l}^C_{35})}{T(\boldsymbol{a}_{1245}, \boldsymbol{l}^B_{12}, \boldsymbol{l}^C_{45}) T(\boldsymbol{a}_{123j}, \boldsymbol{l}^B_{12}, \boldsymbol{l}^C_{3j})} \cdot \frac{\mu_{1245}\mu_{123j}}{\mu_{124j}\mu_{1235}}. \quad (3.32)$$

In this way we can successively compute the projective depths λ_{ij} of the j–points referred to the i–camera. The λ_{ij} will be used in next section for the 3–D reconstruction using the join image concept and the *singular value decomposition* SVD method.

Since this kind of invariant can be also expressed in terms of the quadrifocal tensor [9], we can compute the projective depth based on four cameras.

10.4 Shape and Motion

The orthographic and paraperspective *factorization method for structure and motion* using the affine camera model was developed by Tomasi, Kanade

and Poelman [11, 15]. This method works for cameras viewing small and distance scenes, thus all scale factors of projective depth λ_{ij}=1. For the case of perspective images the scale factors λ_{ij} are unknown. According to Triggs [16] all λ_{ij} satisfy a set of consistency reconstruction equations of the so–called *join image*. One way to compute λ_{ij} is by using the epipolar constraint. If we use a matrix representation this is given by

$$F_{ik}\lambda_{ij}\boldsymbol{x}_{ij} = \boldsymbol{e}_{ik} \wedge \lambda_{kj}\boldsymbol{x}_{kj} \tag{4.33}$$

which after an inner product gives the relation of projective depths for the j-point between camera i and k

$$\lambda'_{kj} = \frac{\lambda_{kj}}{\lambda_{ij}} = \frac{(\boldsymbol{e}_{ik} \wedge \boldsymbol{x}_{kj})F_{ik}\boldsymbol{x}_{ij}}{||\boldsymbol{e}_{ik} \wedge \boldsymbol{x}_{kj}||^2}. \tag{4.34}$$

Considering the i-camera as reference we can norm the λ_{kj} for all k-cameras and use λ'_{kj} instead. If that is not the case we can norm between neighbor images in a chained relationship [16].

In the previous section we presented a better procedure for the computing of λ_{ij} involving three cameras. The extension of the equation (4.34) in terms of the trifocal or quadrifocal tensor is awkward and unpractical.

10.4.1 The join image

The *join image* $\mathcal{J}$ is nothing else than the intersections of optical rays and planes at the points or lines in the 3–D projective space as depicted in figure (10.7). The interrelated geometry can be linearly expressed by the fundamental matrix and trifocal and quadrifocal tensors. The reader will find more details about these linear constraints in chapter 7.

In order to take into account the interrelated geometry, the *projective reconstruction* procedure should put together all the data of the individual images in a geometrically coherent manner. The way to do that is by considering the observations of the points $\boldsymbol{X}_j$ regarding each i–camera

$$\lambda_{ij}\boldsymbol{x}_{ij} = P_i\boldsymbol{X}_j \tag{4.35}$$

as the i–row of a matrix of rank 4. For m cameras and n points the 3m×n matrix $\mathcal{J}$ of the join image is given by

$$\mathcal{J} = \begin{pmatrix} \lambda_{11}\boldsymbol{x}_{11} & \lambda_{12}\boldsymbol{x}_{12} & \lambda_{13}\boldsymbol{x}_{13} & . & . & . & \lambda_{1n}\boldsymbol{x}_{1n} \\ \lambda_{21}\boldsymbol{x}_{21} & \lambda_{22}\boldsymbol{x}_{22} & \lambda_{23}\boldsymbol{x}_{23} & . & . & . & \lambda_{2n}\boldsymbol{x}_{2n} \\ \lambda_{31}\boldsymbol{x}_{31} & \lambda_{32}\boldsymbol{x}_{32} & \lambda_{33}\boldsymbol{x}_{33} & . & . & . & \lambda_{3n}\boldsymbol{x}_{3n} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \lambda_{m1}\boldsymbol{x}_{m1} & \lambda_{m2}\boldsymbol{x}_{m2} & \lambda_{m3}\boldsymbol{x}_{m3} & . & . & . & \lambda_{mn}\boldsymbol{x}_{mn} \end{pmatrix}. \tag{4.36}$$

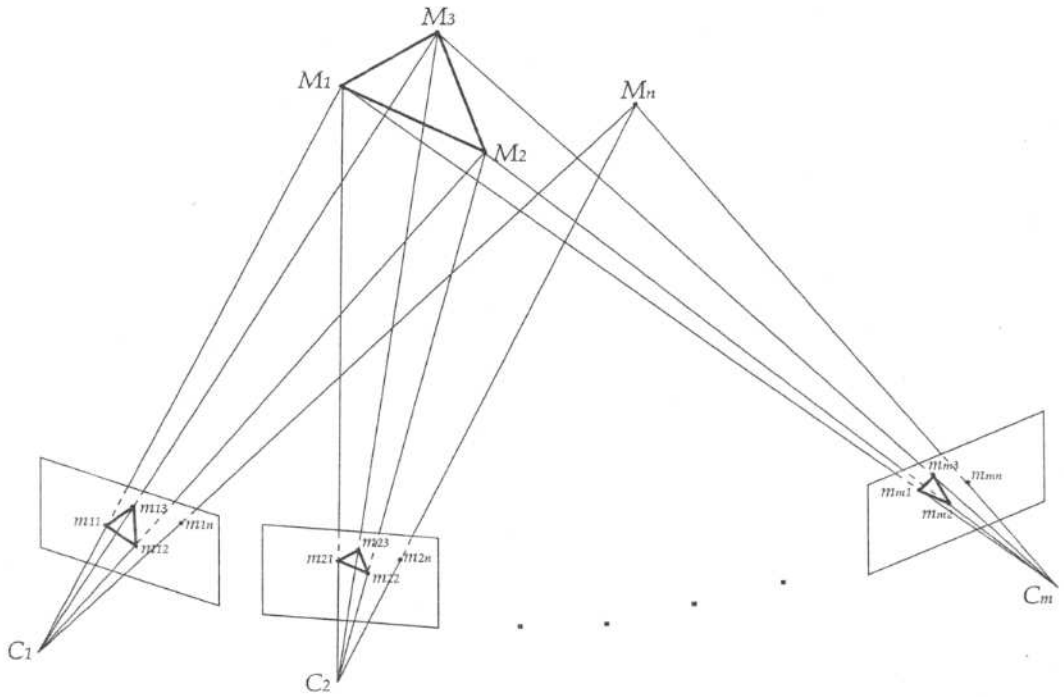


FIGURE 10.7. The geometry of the join image.

For the affine reconstruction procedure the matrix is of rank 3. The matrix $\mathcal{J}$ of the join image is amenable to a singular value decomposition for finding the shape and motion [11, 15].

10.4.2 The SVD method

The application of SVD to $\mathcal{J}$ gives

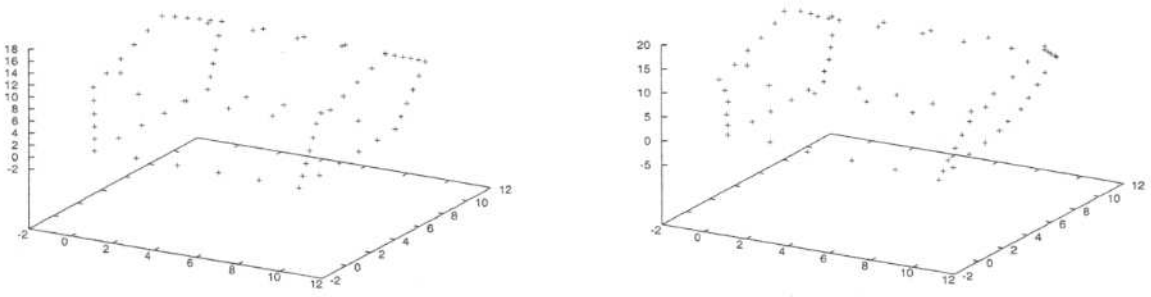
$$\mathcal{J}_{3m\times n} = U_{3m\times r} S_{r\times r} V^T_{n\times r}, \tag{4.37}$$

where the columns of matrix $V^T_{n\times r}$ and $U_{3m\times r}$ constitute the orthonormal base for the input (co-kernel) and output (range) spaces of $\mathcal{J}$. In order to get a decomposition in motion and shape of the projected point structure, $S_{r\times r}$ can be absorbed into both matrices $V^T_{n\times r}$ and $U_{3m\times r}$ as follows

$$\mathcal{J}_{3m\times n} = (U_{3m\times r} S^{\frac{1}{2}}_{r\times r})(S^{\frac{1}{2}}_{r\times r} V^T_{n\times r}) = \begin{pmatrix} P_1 \\ P_2 \\ P_3 \\ . \\ . \\ . \\ P_m \end{pmatrix}_{3m\times 4} (\boldsymbol{X}_1 \boldsymbol{X}_2 \boldsymbol{X}_3 ... \boldsymbol{X}_n)_{4\times n} \tag{4.38}$$

This way to divide $S_{r\times r}$ is not unique. Since the rank of $\mathcal{J}$ is 4 we should take the first four biggest singular values for $S_{r\times r}$. The matrices P_i correspond to the projective mappings or *motion* from the projective space

to the individual images and the point structure or *shape* is given by $\boldsymbol{X}_j$. We test our approach using a simulations program written in Maple. Using the method of section 10.3 firstly we computed the projective depth of the points of a wire house observed with 9 cameras and then using the SVD projective reconstruction method we gained the shape and motion. The reconstructed house after the Euclidean readjustment for the presentation is shown in figure 10.8.



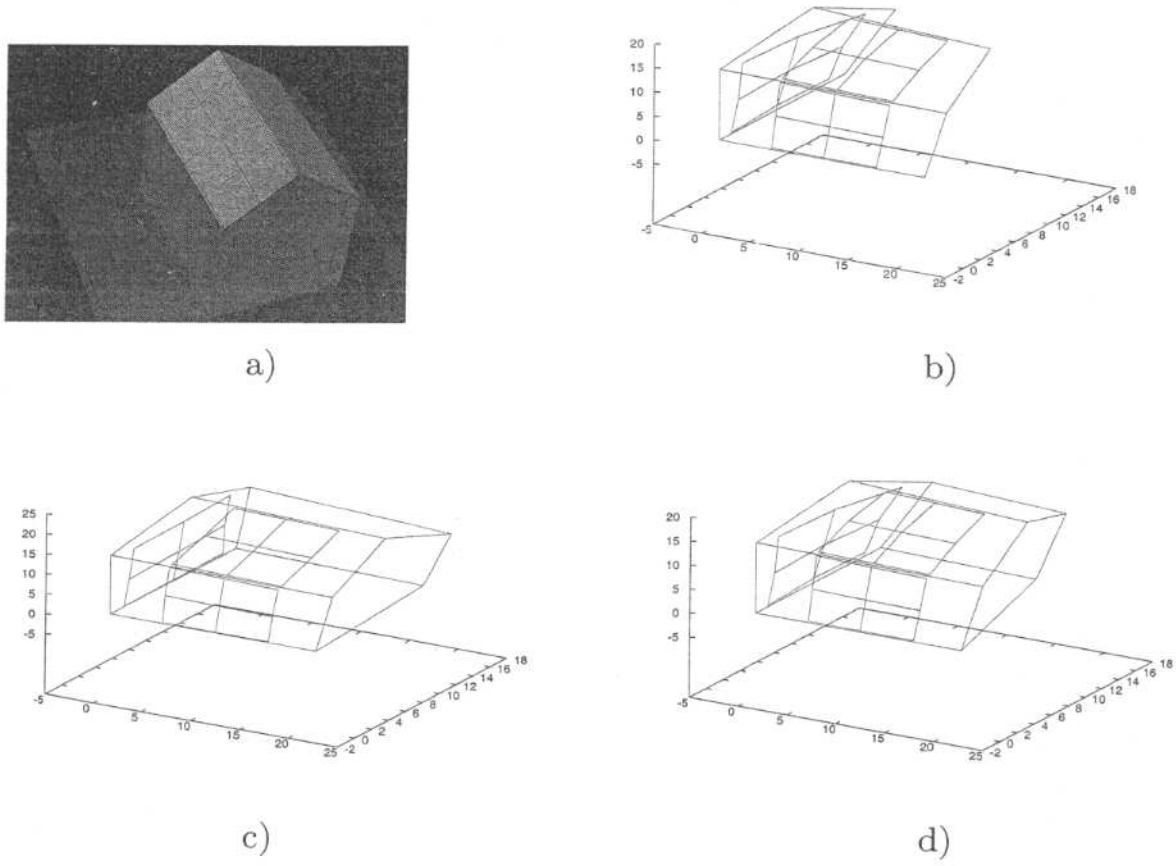
a) b)

FIGURE 10.8. Reconstructed house using a) noise–free observations and b) noisy observations.

We notice that the reconstruction keeps quite well the original form of the model.

The next section will show how we can improve the shape of the reconstructed model using geometric expressions in terms of the operators of algebra of incidence $\vee$ (meet) and $\wedge$ (join) and particular tensor based invariants.

10.4.3 Completion of the 3–D shape using geometric invariants

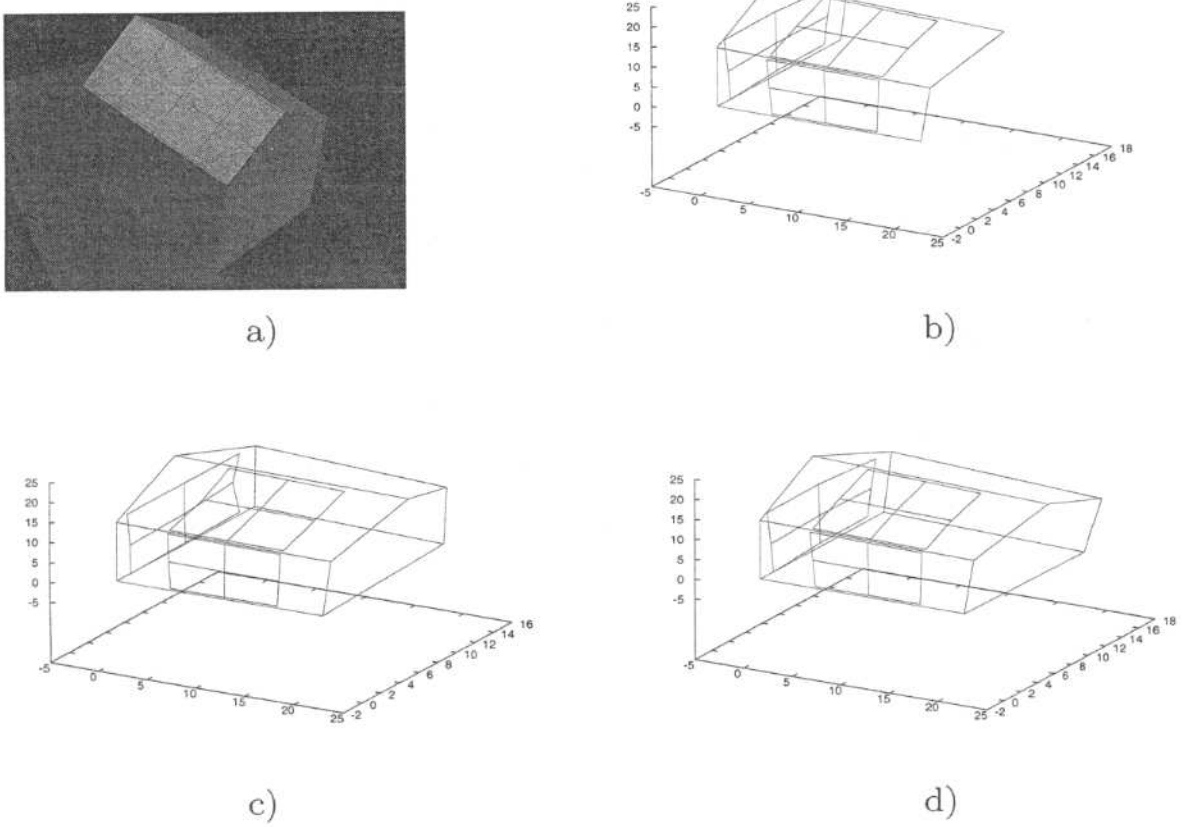


a) b)

c) d)

FIGURE 10.9. a) One of the three images, b) reconstructed incomplete house using 3 images c) extending the join image d) completing in the 3–D space.

The projective structure can be improved in two ways: by completing points on the images, by expanding the join image and then by calling the SVD procedure, or, after the reconstruction, by completing points in the 3–D space like the occluded ones. Both approaches can use geometric inference rules based on symmetries or concrete knowledge about the scene. Using three real views of a similar model house with its most right lower corner missing, see figure 10.9.b , we compute in each image the virtual image point of this 3–D point. Then we reconstruct the scene as shown in figure 10.9.c. As opposite, using geometric incidence operations we completed the house employing the space points as depicted in figure 10.9.d. We can see that creating points in the images yields a better reconstruction of the occluded point. Note that in the reconstructed image we transformed the projective shape into an Euclidean one for the presentation of the results.



a) b)

c) d)

FIGURE 10.10. a) One of the nine images, b) reconstructed incomplete house using 9 images c) extending the join image d) completing in the 3–D space.

We used also lines connecting the reconstructed points only to make visible the house form. Similarly we proceeded using 9 images, as presented in in figure 10.10.a–d.

We can see that the resulting reconstructed point is almost similar in both procedures. As a result we can draw the following conclusion: when we have few views we should extend the join image using virtual image points and in case of several images we should extend the point structure in the 3–D space.

10.5 Conclusions

This chapter focused on the application of projective invariants based on the trifocal tensor. We developed a method to compute the projective depth using this kind of invariants. The resulting projective depths were then used for the initialization of the projective reconstruction of shape and motion.

Furthermore using incidence algebra rules we completed the reconstruction for the case of occluded points.

The main contribution of this paper is that in our geometric method we relate to and extend current approaches regarding projective invariants and their application for reconstruction of shape and motion, as a result the procedures gain geometric transparency and elegance. However, the authors believe that more work have to be done in order to improve the computational algorithms so that the use of projective invariants will be more and more attractive for real systems involving noisy data.

Acknowledgments

Eduardo Bayro-Corrochano was supported by the project SO-201 of the Deutsche Forschungsgemeinschat.