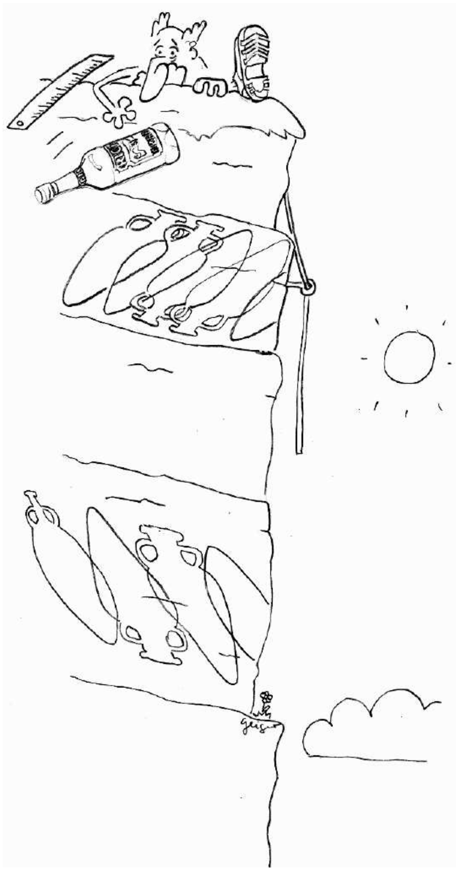
Chapter 11

Stratification of $n \geq 3$ views and applications



Draft-Draft-Draft

11.1 Introduction

This chapter extends chapter 8 to deal with an arbitrary number of cameras. The goal is, as the reader should now be accustomed to, to provide a stratified description of the geometry of the cameras and the structure of the scene according to the information that is available. As in the previous chapters we identify three main levels of description, projective, affine and Euclidean and describe how to compute a canonical representation of the cameras' projection matrices. Many applications are presented along the way to illustrate the theoretical ideas and to show how they can be used to solve real problems.

Section 11.3 is a thorough study of the projective setting. We begin by showing that the description of two views with the Fundamental matrix and of three views with the Trifocal tensors ends with four views. Even though there is such a thing as the Quadrifocal tensors, it is algebraically dependent upon the Fundamental matrices and the Trifocal tensors and therefore probably not worth studying. Moreover, beyond four views there are no further entities and we find that the simplest way to describe the geometry of $n \geq 3$ cameras is through their projection matrices. We thus return to the case of three cameras which will be the building block in the more general case of an arbitrary number of views. We show how to compute a canonical representation of the triplet of projection matrices given either a triplet of Fundamental matrices or a Trifocal tensor. We then discuss the general case and present two main classes of solutions for computing a canonical representation of the n-tuple of projection matrices. The first method, called the incremental one, starts from a canonical representation of an $n-1$-tuple of projection matrices and incorporates an nth view, yielding the desired result. In order to do this, it only requires that the new view shares correspondences with two others. The second method, called the global one, is more restrictive in that it requires that all image features used in the estimation of the canonical representation are seen in all n views. The method is first described in the case of affine cameras and then extended to the general case. Before we leave the subject of the projective stratum, we describe a method for refining the canonical representation which is quite useful not only in this setting but also in the affine and Euclidean frameworks. The method, called the bundle adjustment, is a minimization method that has been developed by photogrammeters and requires a good initialization such as the one produced by the incremental or global methods.

Section 11.4 is a first analysis of the computation of the affine and Euclidean canonical representations of $n \geq 3$ views from the projective and affine canonical representations, respectively. In the affine case the only extra information that is required in order to upgrade the representation from projective to affine is the infinity morphism between any of the n views. In the Euclidean case the only extra information that is required in order to upgrade the representation from affine to Euclidean is the intrinsic parameters of one of the views.

Section 11.5 deals with the case of a sequence of images which are taken by a stereo rig, rather than a single moving camera. The fact that the geometry of the rig is kept fixed opens the door to new possibilities, including the possibility to upgrade to affine representation with a single motion.

11.2 Canonical representations of n views

We now extend the remarks of section 8.1 to the case of an arbitrary number of cameras. Let them be represented by the n-tuple of perspective projection matrixes $(\boldsymbol{\mathcal{P}}_1, \cdots, \boldsymbol{\mathcal{P}}_n)$ in some image and space coordinate systems. A change of coordinates defined by an homography $\boldsymbol{\mathcal{H}}$ changes the perspective projection matrices into $(\boldsymbol{\mathcal{P}}_1\boldsymbol{\mathcal{H}}, \cdots, \boldsymbol{\mathcal{P}}_n\boldsymbol{\mathcal{H}})$. In practice, the homography belongs to one of the three groups $\mathcal{PLG}$, $\mathcal{AG}$ and $\mathcal{E}$, respectively (see chapter 3). For a given n-tuple of projection matrices, we consider its orbit, i.e. all n-tuples obtained by multiplying each projection matrix on the right by an element of the group. All n-tuples in the same orbit are equivalent in the intuitive sense that they all correspond to the same images. Mathematically, each of these three groups acts on n-tuples of projection matrices by multiplication to the right and we are led again to the definition of the following equivalence relation:

$$(\boldsymbol{\mathcal{P}}_1, \cdots, \boldsymbol{\mathcal{P}}_n) \equiv (\boldsymbol{\mathcal{P}}'_1, \cdots, \boldsymbol{\mathcal{P}}'_n) \Longleftrightarrow \exists \boldsymbol{\mathcal{H}} \in \mathcal{G} \quad \text{such that} \quad \boldsymbol{\mathcal{P}}_1 = \boldsymbol{\mathcal{P}}'_1\boldsymbol{\mathcal{H}} \quad \cdots \quad \boldsymbol{\mathcal{P}}_n = \boldsymbol{\mathcal{P}}'_n,$$

where $\mathcal{G}$ is one of the previous three groups. The equivalence classes are the orbits.

As in the case of two cameras, the next step consists in identifying within each orbit an n-tuple which is simple and use it to represent the orbit. We use the same convention as in section 8.1 and choose $\boldsymbol{\mathcal{P}}_1 \simeq \boldsymbol{\mathcal{P}}_0$ when $\mathcal{G} = \mathcal{PLG}$ or $\mathcal{AG}$ and $\boldsymbol{\mathcal{P}}_1 \simeq \mathbf{A}_1\boldsymbol{\mathcal{P}}_0$ when $\mathcal{G} = \mathcal{E}$, where $\mathbf{A}_1$ is the matrix of the internal parameters of the first view, representing a five-parameter affine transformation of $\mathbb{P}^2$. An element of a given orbit of the form $(\boldsymbol{\mathcal{P}}_0, \cdots, \boldsymbol{\mathcal{P}}_n)$ for $\mathcal{G} = \mathcal{PLG}$ or $\mathcal{G} = \mathcal{AG}$ or of the form $(\mathbf{A}_1\boldsymbol{\mathcal{P}}_0, \cdots, \boldsymbol{\mathcal{P}}_n)$ for $\mathcal{G} = \mathcal{E}$ is called a *canonical* element of the orbit. Similarly to the two views case, there exists a four parameter of canonical elements in each orbit when $\mathcal{G} = \mathcal{PLG}$, a one parameter when $\mathcal{G} = \mathcal{AG}$ and a unique element when $\mathcal{G} = \mathcal{E}$.

11.3 Projective stratum

We show that the analysis that started with two views and the concept of the Fundamental matrix, continued with three and the concept of the Trifocal tensor, stops with four views and the Quadrifocal tensor which is completely determined by the Fundamental matrices and the Trifocal tensors.

11.3.1 Beyond the Fundamental matrix and the Trifocal tensor

Let us consider n cameras and their projection matrices $\boldsymbol{\mathcal{P}}_1, \ldots, \boldsymbol{\mathcal{P}}_n$. A 3D point M projects as $m_1, \ldots, m_n$ in the n views. Let $[x_i, y_i, z_i]^T$ be the projective coordinates of pixel m_i. For each camera, we have the equations (5.6):

$$\begin{aligned} x_i \boldsymbol{V}_i^T \mathbf{M} - y_i \boldsymbol{U}_i^T \mathbf{M} = 0 \\ y_i \boldsymbol{W}_i^T \mathbf{M} - z_i \boldsymbol{V}_i^T \mathbf{M} = 0 \\ z_i \boldsymbol{U}_i^T \mathbf{M} - x_i \boldsymbol{W}_i^T \mathbf{M} = 0 \end{aligned}$$

If we concatenate these $3 \times n$ equations and write them in matrix form, we obtain:

$$\boldsymbol{\mathcal{A}} \mathbf{M} = \mathbf{0} \tag{11.1}$$

where the matrix $\boldsymbol{\mathcal{A}}$ is $3 \times n$.

If we consider (11.1) to be a system of linear equations in the unknown vector $\mathbf{M}$, i.e. we assume that we are given the n projection matrices $\boldsymbol{\mathcal{P}}_i$ and the n views m_i of the 3D point M, we see that the nullspace of $\boldsymbol{\mathcal{A}}$ has to be greater than or equal to 1 and therefore its rank has to be less than or equal to 3. Indeed, if the rank was equal to 4, the only solution would be $\mathbf{M} = \mathbf{0}$ which does not represent a projective point.

A consequence of this observation is that all 4×4 minors of $\boldsymbol{\mathcal{A}}$ must be equal to 0. Let us classify those minors. They fall into three categories, those that involve two cameras, those that involve three, and those that involve four. We consider them in turn.

Proposition 11.1 *The 4×4 minors of $\boldsymbol{\mathcal{A}}$ that involve two of the n cameras express the epipolar constraint between the corresponding pixels.*

Proof : Without loss of generality, we consider the minor

$$\begin{vmatrix} x_1\boldsymbol{V}_1^T - y_1\boldsymbol{U}_1^T \\ y_1\boldsymbol{W}_1^T - z_1\boldsymbol{V}_1^T \\ y_2\boldsymbol{W}_2^T - z_2\boldsymbol{V}_2^T \\ z_2\boldsymbol{U}_2^T - x_2\boldsymbol{W}_2^T \end{vmatrix}$$

Because of proposition 4.25 and theorem 4.2 it is equal to the following meet:

$$(x_1\boldsymbol{V}_1 - y_1\boldsymbol{U}_1) \vartriangle (y_1\boldsymbol{W}_1 - z_1\boldsymbol{V}_1) \vartriangle (y_2\boldsymbol{W}_2 - z_2\boldsymbol{V}_2) \vartriangle (z_2\boldsymbol{U}_2 - x_2\boldsymbol{W}_2)$$

We expand the first and the last two meets and obtain

$$y_1 z_2(x_1\boldsymbol{V}_1 \vartriangle \boldsymbol{W}_1 + y_1\boldsymbol{W}_1 \vartriangle \boldsymbol{U}_1 + z_1\boldsymbol{U}_1 \vartriangle \boldsymbol{V}_1) \vartriangle \\ (x_2\boldsymbol{V}_2 \vartriangle \boldsymbol{W}_2 + y_2\boldsymbol{W}_2 \vartriangle \boldsymbol{U}_2 + z_2\boldsymbol{U}_2 \vartriangle \boldsymbol{V}_2)$$

Beside the scale factor $y_1 z_2$, each of the terms in parenthesis is, according to proposition 5.4, the optical ray of the pixels m_1 and m_2, respectively. Therefore, the nullity of the minor is equivalent to the fact that these two optical rays intersect (proposition 4.33) and that is the way we derived the Fundamental matrix in proposition 6.9. □

If we now consider the minors involving three of the cameras, two of the rows must arise from one camera and the other two rows from two others. We have the following proposition

Proposition 11.2 *The* 4×4 *minors of* $\boldsymbol{\mathcal{A}}$ *that involve three of the* n *cameras express Trifocal constraints between the corresponding three pixels.*

Proof : Without loss of generality, we consider the minor

$$\begin{vmatrix} x_1\boldsymbol{V}_1^T - y_1\boldsymbol{U}_1^T \\ y_1\boldsymbol{W}_1^T - z_1\boldsymbol{V}_1^T \\ y_2\boldsymbol{W}_2^T - z_2\boldsymbol{V}_2^T \\ z_3\boldsymbol{U}_3^T - x_3\boldsymbol{W}_2^T \end{vmatrix}$$

Because of proposition 4.25 and theorem 4.2 it is equal to the following meet:

$$(x_1\boldsymbol{V}_1 - y_1\boldsymbol{U}_1) \vartriangle (y_1\boldsymbol{W}_1 - z_1\boldsymbol{V}_1) \vartriangle (y_2\boldsymbol{W}_2 - z_2\boldsymbol{V}_2) \vartriangle (z_3\boldsymbol{U}_3 - x_3\boldsymbol{W}_3)$$

We expand the first two meets and obtain

$$y_1(x_1\boldsymbol{V}_1 \vartriangle \boldsymbol{W}_1 + y_1\boldsymbol{W}_1 \vartriangle \boldsymbol{U}_1 + z_1\boldsymbol{U}_1 \vartriangle \boldsymbol{V}_1) \vartriangle (y_2\boldsymbol{W}_2 - z_2\boldsymbol{V}_2) \vartriangle (z_3\boldsymbol{U}_3 - x_3\boldsymbol{W}_3)$$

and it can be verified that this is equal to

$$y_1 \mathbf{m}_1^T \boldsymbol{\mathcal{T}}_1(\mathbf{m}_2 \times \varepsilon_1, \mathbf{m}_3 \times \varepsilon_2)$$

Beside the scale factor y_1, this is one of the trilinearities described in section 10.1.2.2 based on the Trifocal tensor. □

If we now consider the minors involving four of the cameras, each row has to arise from one different camera, and we have the following proposition

Proposition 11.3 *The 4×4 minors of $\boldsymbol{\mathcal{A}}$ that involve four of the n cameras express the Quadrifocal constraints between the corresponding four pixels. These constraints can be simply expressed as functions of the Trifocal and epipolar constraints.*

Proof : Without loss of generality, we consider the minor

$$Q = \begin{vmatrix} x_1 \boldsymbol{V}_1^T - y_1 \boldsymbol{U}_1^T \\ x_2 \boldsymbol{V}_2^T - y_2 \boldsymbol{U}_2^T \\ x_3 \boldsymbol{V}_3^T - y_3 \boldsymbol{U}_3^T \\ x_4 \boldsymbol{V}_4^T - y_4 \boldsymbol{U}_4^T \end{vmatrix}$$

Because of proposition 4.25 and theorem 4.2 it is equal to the following meet:

$$(x_1 \boldsymbol{V}_1 - y_1 \boldsymbol{U}_1) \vartriangle (x_2 \boldsymbol{V}_2 - y_2 \boldsymbol{U}_2) \vartriangle (x_3 \boldsymbol{V}_3 - y_3 \boldsymbol{U}_3) \vartriangle (x_4 \boldsymbol{V}_4 - y_4 \boldsymbol{U}_4)$$

In order to show that this is algebraically dependent of quantities involving Fundamental matrixes and Trifocal tensors, we use the Cramer relation (proposition 4.23). We note $\mathbf{\Pi}_i$, $i = 1, 2, 3, 4$ the vector $x_i \boldsymbol{V}_i - y_i \boldsymbol{U}_i$, and $\mathbf{\Pi}$ the vector $y_1 \boldsymbol{W}_1 - z_1 \boldsymbol{V}_1$. These five vectors are related by the Cramer relation

$$\begin{aligned} \mid \mathbf{\Pi}, \mathbf{\Pi}_2, \mathbf{\Pi}_3, \mathbf{\Pi}_4 \mid \mathbf{\Pi}_1 - \mid \mathbf{\Pi}_1, \mathbf{\Pi}_2, \mathbf{\Pi}_3, \mathbf{\Pi}_4 \mid \mathbf{\Pi} + \\ \mid \mathbf{\Pi}_1, \mathbf{\Pi}, \mathbf{\Pi}_3, \mathbf{\Pi}_4 \mid \mathbf{\Pi}_2 - \mid \mathbf{\Pi}_1, \mathbf{\Pi}, \mathbf{\Pi}_2, \mathbf{\Pi}_4 \mid \mathbf{\Pi}_3 + \\ \mid \mathbf{\Pi}_1, \mathbf{\Pi}, \mathbf{\Pi}_2, \mathbf{\Pi}_3 \mid \mathbf{\Pi}_4 = \mathbf{0} \end{aligned}$$

We take the meet of both sides of this equation on the right with $\mathbf{\Pi}_1 \vartriangle \boldsymbol{U}_2 \vartriangle \boldsymbol{V}_2$. This eliminates the first and third term of the previous sum. We obtain

$$\begin{aligned} Q \mid \mathbf{\Pi}_1, \mathbf{\Pi}, \boldsymbol{U}_2, \boldsymbol{V}_2 \mid + \mid \mathbf{\Pi}_1, \mathbf{\Pi}, \mathbf{\Pi}_2, \mathbf{\Pi}_4 \mid\mid \mathbf{\Pi}_1, \mathbf{\Pi}_3, \boldsymbol{U}_2, \boldsymbol{V}_2 \mid - \\ \mid \mathbf{\Pi}_1, \mathbf{\Pi}, \mathbf{\Pi}_2, \mathbf{\Pi}_3 \mid\mid \mathbf{\Pi}_1, \mathbf{\Pi}_4, \boldsymbol{U}_2, \boldsymbol{V}_2 \mid = \mathbf{0} \end{aligned}$$

The coefficient of Q is equal to $y_1(\mathbf{F}_{12}\mathbf{m}_1)_3$ (proposition 6.9). The other two terms involve only the Trifocal tensors. In detail $|\ \mathbf{\Pi}_1, \mathbf{\Pi}, \mathbf{\Pi}_2, \mathbf{\Pi}_4\ |$ and and $|\ \mathbf{\Pi}_1, \mathbf{\Pi}_4, \boldsymbol{U}_2, \boldsymbol{V}_2\ |$ involve two of the Trifocal tensors between views 1, 2 and 4, $|\ \mathbf{\Pi}_1, \mathbf{\Pi}_3, \boldsymbol{U}_2, \boldsymbol{V}_2\ |$ and $|\ \mathbf{\Pi}_1, \mathbf{\Pi}, \mathbf{\Pi}_2, \mathbf{\Pi}_3\ |$ two of the Trifocal tensors between views 1, 2 and 3. □

Proposition 11.3 shows that Quadrifocal tensors do exist but do not add more information than the Fundamental matrices and the Trifocal tensors. Therefore, when we consider more than three views it is unnecessary to attempt to generalize the ideas of the Fundamental matrices and Trifocal tensors: there is such a generalization for four views but it is completely determined by the the lower order tensors and there is no such thing for more than four views.

11.3.2 The projection matrices: three views

According to the previous section, it is unnecessary to worry about higher order relations between views than those provided by the binocular and trinocular constraints that have been studied in chapters 6 and 9. Therefore, in the case of four and more views we will use only the binocular and trinocular constraints to compute the projection matrices of the views.

We start with the case of three views which we will then generalize to an arbitrary number of views. Suppose that we have estimated all three Fundamental matrices and that they are compatible in the sense of chapter 9, definition 9.1. The question that we address again is what kind of 3D information can we extract about the world. Just like in the case of three views analysed in chapter 8, the answer turns out to be that only projective information can be recovered from the Fundamental matrices. We show this explicitely by computing the canonical representation of the triple $(\boldsymbol{\mathcal{P}}_1, \boldsymbol{\mathcal{P}}_2, \boldsymbol{\mathcal{P}}_3)$. Similarly to the two views case, there is a four-parameter family of matrices $(\boldsymbol{\mathcal{P}}_2, \boldsymbol{\mathcal{P}}_3)$ such that that $\boldsymbol{\mathcal{P}}_1 \simeq \boldsymbol{\mathcal{P}}_0$ and the three Fundamental matrices obtained from those three perspective projection matrices are proportional to the given Fundamental matrices:

Proposition 11.4 *For a given triple of compatible matrices* $(\mathbf{F}_{12}, \mathbf{F}_{23}, \mathbf{F}_{31})$ *(see definition 9.1), the condition* $\boldsymbol{\mathcal{P}} \simeq \boldsymbol{\mathcal{P}}_0$ *implies that* $\boldsymbol{\mathcal{P}}_2$ *and* $\boldsymbol{\mathcal{P}}_3$ *take the simple forms:*

$$\boldsymbol{\mathcal{P}}_2 = \left[\mathbf{S}_{12} + \mathbf{e}_{21}\mathbf{r}_2^T\ \mu_2\mathbf{e}_{21}\right] = [\mathbf{S}_{12}, \mathbf{e}_{21}] \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{r}_2^T & \mu_2 \end{bmatrix} \tag{11.2}$$

and:

$$\mathcal{P}_3 = \left[\mathbf{S}_{13} + \mathbf{e}_{31}\mathbf{r}_3^T\ \mu_3\mathbf{e}_{31}\right] = [\mathbf{S}_{13}, \mathbf{e}_{31}] \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{r}_3^T & \mu_3 \end{bmatrix} \tag{11.3}$$

The values of μ_2 and μ_3 satisfy the following relation:

$$\mu_2 : \mu_3 = -\mathbf{S}_{12}^T\mathbf{F}_{23}^T\mathbf{e}_{31} : \mathbf{S}_{13}^T\mathbf{F}_{23}\mathbf{e}_{21},$$

the vectors appearing on the righthand side being parallel, representing the Trifocal line t_1 in the first image. The vectors $\mathbf{r}_2$ and $\mathbf{r}_3$ are related as follows:

$$\mathbf{r}_2 + \alpha\mathbf{r}_3 = \mathbf{r}$$

where $\alpha = -\mu_2/\mu_3$ and $\mathbf{r}$ a vector which can be computed from the the three Fundamental matrices and the epipoles as explained in the proof.

Proof : Equations (11.2) and (11.3) are immediate consequences of proposition 11.4. $\mathbf{r}_2$ (respectively $\mathbf{r}_3$) is the image in view 1 of the line of intersection of an arbitrary plane Π_1 (respectively Π_2) not going through the optical center C_1 with the intrinsic plane of the second camera with respect to the first view, i.e. determined by the line represented by $\mathbf{e}_{21}$ (respectively with the intrinsic plane of the third camera with respect to the first view, i.e. determined by the line represented by $\mathbf{e}_{31}$). We have so far only taken into account the Fundamental matrices $\mathbf{F}_{12}$ and $\mathbf{F}_{13}$. The consideration of $\mathbf{F}_{23}$ allows us to show that if $\mathbf{r}_2$ and μ_2 are given, $\mathbf{r}_3$ and μ_3 are fixed.

Consider a projective point P represented by the vector $\left[\mathbf{p}^T,\ q\right]$ in the current projective frame. Its image p_2 in view 2 is represented by:

$$\mathbf{p}_2 \simeq \mathcal{P}_2\mathbf{P} \simeq \mathbf{S}_{12}\mathbf{p} + (\mathbf{r}_2^T\mathbf{p} + \mu_2 q)\mathbf{e}_{21}$$

Similarly, its image p_3 in view 3 is represented by:

$$\mathbf{p}_3 \simeq \mathcal{P}_3\mathbf{P} \simeq \mathbf{S}_{13}\mathbf{p} + (\mathbf{r}_3^T\mathbf{p} + \mu_3 q)\mathbf{e}_{31}$$

We now express the fact that $\mathbf{p}_3^T\mathbf{F}_{23}\mathbf{p}_2$ must be 0 for all points P. We use the fact that since the three Fundamental matrices are compatible, $\mathbf{e}_{31}^T\mathbf{F}_{23}\mathbf{e}_{21} = 0$, and obtain $\mathbf{p}_3^T\mathbf{F}_{23}\mathbf{p}_2 = \mathbf{P}^T\mathbf{A}\mathbf{P}$.

The 4×4 matrix $\mathcal{A}$ is given by:

$$\mathcal{A} \simeq \begin{bmatrix} \mathbf{B} + \mathbf{r}_2\mathbf{u}^T + \mathbf{r}_3\mathbf{v}^T & \mathbf{0}_3 \\ \mu_2\mathbf{u}^T + \mu_3\mathbf{v}^T & 0 \end{bmatrix},$$

where the matrix $\mathbf{B} = \mathbf{S}_{13}^T\mathbf{F}_{23}\mathbf{S}_{12}$, the vector $\mathbf{u}^T = \mathbf{e}_{21}^T\mathbf{F}_{23}^T\mathbf{S}_{13}$ and the vector $\mathbf{v}^T = \mathbf{e}_{31}^T\mathbf{F}_{23}\mathbf{S}_{12}$. The condition $\mathbf{P}^T\mathbf{A}\mathbf{P} = 0$ for all $\mathbf{P}$ is therefore equivalent to the two conditions:

$$\mu_2\mathbf{u} + \mu_3\mathbf{v} = \mathbf{0},$$

and the matrix $\mathbf{C} = \mathbf{B} + \mathbf{r}_2\mathbf{u}^T + \mathbf{r}_3\mathbf{v}^T$ is antisymmetric.

The reader will easily convince himself that the two vectors $\mathbf{u}$ and $\mathbf{v}$ are two representations of the Trifocal line t_1 in the first image. Let us verify this for $\mathbf{u}$. From chapter 9, section 9.1.1, we know that the vector $\mathbf{F}_{23}\mathbf{e}_{21}$ represents the Trifocal line t_3. The matrix $\mathbf{S}_{13}^T \simeq \mathbf{F}_{13}^T[\mathbf{e}_{31}]_\times = \mathbf{F}_{31}[\mathbf{e}_{31}]_\times$ represents a planar morphism from the set of lines in image 3 to the set of lines in image 1. Applying it to the Trifocal line t_3 is therefore equivalent to applying $\mathbf{F}_{31}$ to the point of intersection of t_3 and the line represented by the vector $\mathbf{e}_{31}$. Since this is a point of t_3, when we apply $\mathbf{F}_{31}$ to it we obtain its epipolar line in the first image, i.e. t_1.

The first of the two conditions therefore determines $\mu_2 : \mu_3$ as stated in the proposition.

Let us now write $\mathbf{v} = \alpha\mathbf{u}$, where $\alpha = -\mu_2/\mu_3$ is known. Using the definitions $\mathbf{S}_{ij} = [\mathbf{e}_{ji}]_\times \mathbf{F}_{ij}$ (we can scale the three fudamental matrices so that this is true), the second condition $\mathbf{C} + \mathbf{C}^T$ can be rewritten as:

$$
\begin{aligned}
&(\mathbf{r}_2 + \alpha\mathbf{r}_3)\mathbf{u}^T + \mathbf{u}(\mathbf{r}_2 + \alpha\mathbf{r}_3)^T = \\
&\qquad\qquad \mathbf{F}_{13}^T [\mathbf{e}_{31}]_\times \mathbf{F}_{23} [\mathbf{e}_{21}]_\times \mathbf{F}_{12} + \mathbf{F}_{12}^T [\mathbf{e}_{21}]_\times \mathbf{F}_{23}^T [\mathbf{e}_{31}]_\times \mathbf{F}_{13} \equiv \mathbf{D}
\end{aligned}
$$

We note that the lefthand side is a matrix of rank 2, and that its nullspace is generated by the vector $(\mathbf{r}_2 + \alpha\mathbf{r}_3) \times \mathbf{u}$. Similarly, the righthand side is also of rank two since one verifies that $\mathbf{D}\mathbf{e}_{12} \simeq \mathbf{D}\mathbf{e}_{13}$, each vector being a representation of the Trifocal line t_1. Let us do it for example for $\mathbf{D}\mathbf{e}_{12}$. Up to the sign, this is the sum of $\mathbf{B}\mathbf{e}_{12}$, which is equal to $\mathbf{0}$ since $\mathbf{S}_{12}\mathbf{e}_{12} = \mathbf{0}$ and, $\mathbf{B}^T\mathbf{e}_{12} = \mathbf{S}_{12}^T\mathbf{F}_{23}^T\mathbf{S}_{13}\mathbf{e}_{12}$. According to the previous discussion, $\mathbf{S}_{13}\mathbf{e}_{12} = [\mathbf{e}_{31}]_\times\mathbf{F}_{13}\mathbf{e}_{12}$ represents a point on the Trifocal line t_3 which is mapped by $\mathbf{F}_{23}^T = \mathbf{F}_{32}$ to the Trifocal line t_2 which is mapped by $\mathbf{S}_{12}^T$ to the Trifocal line t_1.

We can therefore compute two numbers β and γ such that $\mathbf{D}(\beta\mathbf{e}_{12} + \gamma\mathbf{e}_{13}) = \mathbf{0}$.

There is a slight difficulty here with respect to the choice of scale for the two vectors $\mathbf{e}_{12}$ and $\mathbf{e}_{13}$. We know from our discussions in chapter 8 that in order to write $\boldsymbol{\mathcal{P}}_2$ and $\boldsymbol{\mathcal{P}}_3$ as in equations (11.2) and (11.3), respectively, we fixed the scale of the vectors $\mathbf{e}_{21}$ and $\mathbf{e}_{31}$ but *not* of $\mathbf{e}_{12}$ and $\mathbf{e}_{13}$. In order to compute β and γ, we need to fix their scale. The simplest is to assume that they have unit norm, the final result does not depend, of course, on this choice.

Therefore we must have $\beta\mathbf{e}_{12} + \gamma\mathbf{e}_{13} \simeq (\mathbf{r}_2 + \alpha\mathbf{r}_3) \times \mathbf{u}$. Let us note $\mathbf{s}$ the known vector $\beta\mathbf{e}_{12} + \gamma\mathbf{e}_{13}$; it is orthogonal to $\mathbf{u}$ since the epipoles e_{12} and e_{13} are on the Trifocal line t_1 represented by $\mathbf{u}$. This implies that the vector

$\mathbf{r} = \mathbf{r}_2 + \alpha \mathbf{r}_3$ is a linear combination of $\mathbf{u}$ and $\mathbf{u} \times \mathbf{s}$:

$$\mathbf{r} = a\mathbf{u} + b\mathbf{u} \times \mathbf{s}$$

Rewriting the constraint $\mathbf{C} + \mathbf{C}^T = \mathbf{0}$ as a function of the unknowns a, b, the known vectors $\mathbf{u}$, $\mathbf{s}$ and the known matrix $\mathbf{D}$, we obtain:

$$2a\mathbf{u}\mathbf{u}^T + b\left[(\mathbf{u} \times \mathbf{s})\mathbf{u}^T + \mathbf{u}(\mathbf{u} \times \mathbf{s})^T\right] = \mathbf{D}$$

Applying both sides of the equation to the known vector $\mathbf{u}$, we obtain a vector equation that yields the values of a and b:

$$(2a\mathbf{u} + b(\mathbf{u} \times \mathbf{s}))\|\mathbf{u}\|^2 = \mathbf{D}\mathbf{u}$$

The explicit values are

$$a = \frac{1}{2}\frac{\mathbf{u}^T\mathbf{D}\mathbf{u}}{\|\mathbf{u}\|^4}$$
$$b = \frac{(\mathbf{u} \times \mathbf{s})^T\mathbf{D}\mathbf{u}}{\|\mathbf{u} \times \mathbf{s}\|^2\|\mathbf{u}\|^2}$$

From the values of a and b and the relation $\mathbf{r}_2 + \alpha\mathbf{r}_3 = \mathbf{r} = a\mathbf{u} + b\mathbf{u} \times \mathbf{s}$ we conclude that, given $\mathbf{r}_2$, we can compute $\mathbf{r}_3$.

Note that the proof gives an explicit way of computing the quantities $(\mathbf{r}_3, \mu_3)$ as functions of $(\mathbf{r}_2, \mu_2)$ from the three Fundamental matrices, the two epipoles (e_{12}, e_{13}) in the first image, the epipole e_{21} in the second image and the epipole e_{31} in the third image. □

Unfortunately, the previous method for computing $\boldsymbol{\mathcal{P}}_2$ and $\boldsymbol{\mathcal{P}}_3$ fails if, for example, the three optical centers are aligned since in that case the Trifocal lines are not defined. Therefore we need another method for that case. Moreover, if we have estimated the Trifocal tensor between the three views using the methods described in chapter 10, we would like to use it directly to compute the perspective projection matrices without going through the Fundamental matrices. We are going to show that if we use the Trifocal tensor which, as seen in chapter 9, does not induce degeneracies in the case of aligned optical centers, we do not run into any problems. The method developed will also provide a direct way of computing the perspective projection matrices from the Trifocal tensor.

Proposition 11.5 *For a given Trifocal tensor $\boldsymbol{\mathcal{T}}$, the condition $\boldsymbol{\mathcal{P}}_1 \simeq \boldsymbol{\mathcal{P}}_0$ implies that $\boldsymbol{\mathcal{P}}_2$ and $\boldsymbol{\mathcal{P}}_3$ take the simple forms (11.2) and (11.3). The ratio* *of μ_2 and μ_3 satisfies the following relation:*

$$\mu_2 : \mu_3 = \mid \boldsymbol{\mathcal{T}}(\mathbf{l}_2,\, \mathbf{l}_3),\, \boldsymbol{\mathcal{T}}(\mathbf{e}_{21}, \mathbf{e}_{31}),\, \mathbf{S}_{12}^T \mathbf{l}_2 \mid \cdot (\mathbf{l}_3^T \mathbf{e}_{31}) : \\ \mid \boldsymbol{\mathcal{T}}(\mathbf{l}_2,\, \mathbf{l}_3),\, \boldsymbol{\mathcal{T}}(\mathbf{e}_{21}, \mathbf{e}_{31}),\, \mathbf{S}_{13}^T \mathbf{l}_3 \mid \cdot (\mathbf{l}_2^T \mathbf{e}_{21}) \quad (11.4)$$

for "almost all" pairs $(l_2,\, l_3)$ of lines in images 2 and 3, respectively. What we mean by "almost all" is detailed in the proof. The vectors $\mathbf{r}_2$ and $\mathbf{r}_3$ are related as follows:

$$\alpha \mathbf{r}_3 - \mathbf{r}_2 = \mathbf{r} \quad (11.5)$$

where $\alpha = \frac{\mu_2}{\mu_3}$ and $\mathbf{r}$ is given by:

$$\mathbf{r} = \beta \frac{\boldsymbol{\mathcal{T}}(\mathbf{e}_{21}, \mathbf{e}_{31})}{\|\mathbf{e}_{21}\|^2 \cdot \|\mathbf{e}_{31}\|^2} \quad (11.6)$$

where:

$$\beta = \frac{\mid \mathbf{S}_{12}^T \mathbf{l}_2,\, \mathbf{S}_{13}^T \mathbf{l}_3,\, \boldsymbol{\mathcal{T}}(\mathbf{e}_{21}, \mathbf{e}_{31}) \mid \cdot (\mathbf{l}_3^T \mathbf{e}_{31})}{\mid \mathbf{S}_{13}^T \mathbf{l}_3,\, \boldsymbol{\mathcal{T}}(\mathbf{l}_2,\, \mathbf{l}_3),\, \boldsymbol{\mathcal{T}}(\mathbf{e}_{21}, \mathbf{e}_{31}) \mid} \quad (11.7)$$

It is always possible to find two lines l_2 and l_3 such that equations (11.4) and (11.7) are well-defined and β is nonzero.

Proof : We know from chapter 9 that the two Fundamental matrices $\mathbf{F}_{12}$ and $\mathbf{F}_{13}$ are easily computed from $\boldsymbol{\mathcal{T}}$, hence equations (11.2) and (11.3). Instead of using $\mathbf{F}_{23}$ as in the proof of proposition 11.4, we use the Trifocal tensor. From equations (11.2) and (11.3), using equations (9.12), (9.13) and (9.14), we can write the matrices $\mathbf{G}^n$ of $\boldsymbol{\mathcal{T}}$ as functions of the matrices $\mathbf{S}_{12}$, $\mathbf{S}_{13}$, the vectors $\mathbf{r}_2$ and $\mathbf{r}_3$ and the two numbers μ_2 and μ_3:

$$\mathbf{G}^n \simeq \mu_2 \mathbf{e}_{21} \mathbf{S}_{13}^{(n)^T} - \mu_3 \mathbf{S}_{12}^{(n)} \mathbf{e}_{31}^T + (\mu_2 r_3^n - \mu_3 r_2^n) \mathbf{e}_{21} \mathbf{e}_{31}^T \; n = 1,2,3$$

where r_2^n and r_3^n are the nth coordinates of the vectors $\mathbf{r}_2$ and $\mathbf{r}_3$. It is important to note that we have written "$\simeq$" and not "$=$". This is because the given Trifocal tensor $\boldsymbol{\mathcal{T}}$ is only proportional, not equal, to the one computed from the three perspective projection matrices. Note that the scale factor is global, i.e. it is the same for the three matrices $\mathbf{G}^n$. Since the right-hand side of the previous equation is homogeneous in $(\mu_2,\, \mu_3)$, the scale factor can be written $1/(a\mu_2 + b\mu_3)$, where a and b are unknown:

$$\mathbf{G}^n = \frac{1}{a\mu_2 + b\mu_3} (\mu_2 \mathbf{e}_{21} \mathbf{S}_{13}^{(n)^T} - \mu_3 \mathbf{S}_{12}^{(n)} \mathbf{e}_{31}^T + (\mu_2 r_3^n - \mu_3 r_2^n) \mathbf{e}_{21} \mathbf{e}_{31}^T) \\ n = 1,2,3 \quad (11.8)$$

Note that we have necessarily $a\mu_2 + b\mu_3 \neq 0$.

Let us now compute $\mathcal{T}(\mathbf{e}_{21},\, \mathbf{e}_{31})$ using those expressions. Since ${\mathbf{S}_{13}^{(n)}}^T \mathbf{e}_{31} = {\mathbf{S}_{12}^{(n)}}^T \mathbf{e}_{21} = 0$ $n = 1, 2, 3$, we have:

$$\frac{\mathcal{T}(\mathbf{e}_{21}, \mathbf{e}_{31})}{\|\mathbf{e}_{21}\|^2 \cdot \|\mathbf{e}_{31}\|^2} = \frac{\mu_2 \mathbf{r}_3 - \mu_3 \mathbf{r}_2}{a\mu_2 + b\mu_3}$$

Therefore

$$\frac{\mu_2}{\mu_3}\mathbf{r}_3 - \mathbf{r}_2 = (b + a\frac{\mu_2}{\mu_3})\frac{\mathcal{T}(\mathbf{e}_{21}, \mathbf{e}_{31})}{\|\mathbf{e}_{21}\|^2 \cdot \|\mathbf{e}_{31}\|^2}$$

Let $\alpha = \frac{\mu_2}{\mu_3}$ and $\beta = a\alpha + b$, an rewrite the previous equation:

$$\alpha \mathbf{r}_3 - \mathbf{r}_2 = \beta \frac{\mathcal{T}(\mathbf{e}_{21}, \mathbf{e}_{31})}{\|\mathbf{e}_{21}\|^2 \cdot \|\mathbf{e}_{31}\|^2}$$

Let us now consider an arbitrary pair $(l_2,\, l_3)$ of lines in images 2 and 3, respectively. Using the previous equation, we can eliminate the term $\frac{\mu_2 \mathbf{r}_3 - \mu_3 \mathbf{r}_2}{a\mu_2 + b\mu_3}$ from equation (11.8) and write:

$$\mathcal{T}(\mathbf{l}_2,\, \mathbf{l}_3) = \frac{\mathcal{T}(\mathbf{e}_{21}, \mathbf{e}_{31})}{\|\mathbf{e}_{21}\|^2 \cdot \|\mathbf{e}_{31}\|^2}(\mathbf{e}_{21}^T \mathbf{l}_2)(\mathbf{e}_{31}^T \mathbf{l}_3) + \frac{1}{a\mu_2 + b\mu_3}(\mu_2(\mathbf{e}_{21}^T \mathbf{l}_2)\mathbf{S}_{13}^T \mathbf{l}_3 - \mu_3(\mathbf{e}_{31}^T \mathbf{l}_3)\mathbf{S}_{12}^T \mathbf{l}_2)$$

We then note that for each point m of the line l_{23} represented by $\mathcal{T}(\mathbf{e}_{21}, \mathbf{e}_{31})$ we have:

$$\mathbf{m}^T \mathcal{T}(\mathbf{l}_2,\, \mathbf{l}_3) = \frac{1}{a\mu_2 + b\mu_3}(\mu_2(\mathbf{e}_{21}^T \mathbf{l}_2)(\mathbf{m}^T \mathbf{S}_{13}^T \mathbf{l}_3) - \mu_3(\mathbf{e}_{31}^T \mathbf{l}_3)(\mathbf{m}^T \mathbf{S}_{12}^T \mathbf{l}_2)) \tag{11.9}$$

Choosing two arbitrary points on l_{23} yields α. Let us for example choose the following two points: m_3 is represented by $\mathbf{S}_{13}^T \mathbf{l}_3 \times \mathbf{l}_{23}$ and m_2 by $\mathbf{S}_{12}^T \mathbf{l}_2 \times \mathbf{l}_{23}$. Taking $m = m_2$, $m = m_3$ in equation (11.9), and taking the ratio to eliminate $a\mu_2 + b\mu_3$, we obtain the ratio $\mu_2 : \mu_3$:

$$\frac{\mathbf{m}_2^T \mathcal{T}(\mathbf{l}_2,\, \mathbf{l}_3)}{\mathbf{m}_3^T \mathcal{T}(\mathbf{l}_2,\, \mathbf{l}_3)} = -\frac{\mu_2}{\mu_3}\frac{(\mathbf{l}_2^T \mathbf{e}_{21})(\mathbf{m}_2^T \mathbf{S}_{13}^T \mathbf{l}_3)}{(\mathbf{l}_3^T \mathbf{e}_{31})(\mathbf{m}_3^T \mathbf{S}_{12}^T \mathbf{l}_2)}$$

We notice that $\mathbf{m}_3^T \mathbf{S}_{12}^T \mathbf{l}_2 = -\mathbf{m}_2^T \mathbf{S}_{13}^T \mathbf{l}_3$ and obtain equation (11.4).

Having obtained $\alpha = \mu_2/\mu_3$, it is easy to obtain $\beta = a\alpha + b$ from equation (11.9): choosing $m = m_3$ we obtain equation (11.7).

Since both μ_2 and μ_3 are nonzero, the method fails when any of the factors in equations (11.4) and (11.7) become 0. The question thus is, can we choose the two lines l_2 and l_3 so that this does not happen?

The factor $\mathbf{l}_2^T\mathbf{e}_{21}$ (respectively $\mathbf{l}_3^T\mathbf{e}_{31}$) becomes 0 iff l_2 (respectively l_3) is an epipolar line for the pair (1, 2) (respectively (1, 3)). The factor $\mathbf{S}_{12}^T\mathbf{l}_2$ (respectively $\mathbf{S}_{13}^T\mathbf{l}_3$) is zero iff $l_2 = e_{21}$ (respectively $l_3 = e_{31}$). The vector $\boldsymbol{\mathcal{T}}(\mathbf{e}_{21}, \mathbf{e}_{31})$ is never zero, otherwise it would imply that the line of intersection of the two planes (C_2, e_{21}) and (C_3, e_{31}) would go through C_1 and therefore that those two planes would be epipolar planes for the pairs (1, 2) and (1, 3), respectively. But this is impossible since the two lines e_{21} and e_{31} do not, by construction, go through the epipoles e_{21} and e_{31}. Similarly, the line l_{23} does not go either through the epipoles e_{12} or e_{13}. To summarize, we have so far discarded the following possibilities:

- l_2 (respectively l_3) is not an epipolar line for the pair (1, 2) (respectively (1, 3)) of views.
- l_2 (respectively l_3) is different from the line e_{21} (respectively e_{31}).

According to theorem 9.1, the term $\boldsymbol{\mathcal{T}}(\mathbf{l}_2, \mathbf{l}_3)$ is equal to $\mathbf{0}$ if and only if the pair (l_2, l_3) is a pair of corresponding epipolar lines for the pair (2, 3) of views or a pair of non corresponding epipolar lines if the three optical centers are aligned. Therefore we must also discard those possibilities.

Having done this, the only remaining requirements are that both determinants in the numerator and denominator of (11.4) are nonzero. Geometrically this is equivalent to choosing l_2 and l_3 in such a way that $\boldsymbol{\mathcal{T}}(\mathbf{l}_2, \mathbf{l}_3)$ does not go through either m_2 or m_3 and this is always possible. A feeling for this can be obtained by looking at figure 11.1. $\square$

11.3.3 The projection matrices: arbitrary number of views

We now deal with the general problem of computing the projective canonical representations of an arbitrary number n of views from a set of Fundamental matrices and Trifocal tensors. We present two techniques for achieving this. The first method is sequential and does not require all feature points to be seen in all views. The second method is global and requires that all feature points are seen in all views.

11.3.3.1 The incremental method

We have seen in the previous section how to compute a four-parameter family of perspective projection matrices for three cameras given their three

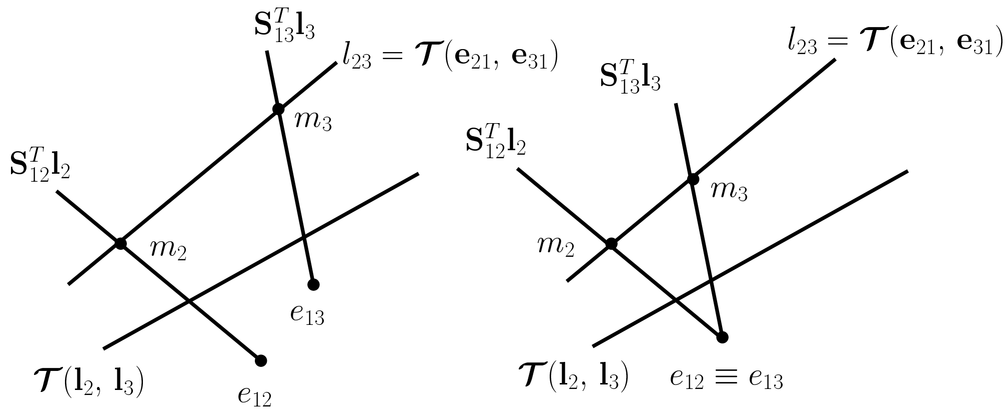


Figure 11.1: The choice of the point m on the line l_{23} (see text).

Fundamental matrices or one of their Trifocal tensor. This can be extended to an arbitrary number n, $n > 3$.

Suppose we have a canonical representation of $n-1$ views. What this means is that we have chosen one of the $n-1$ views as the reference (we will assume without loss of generality that it is view 1) and a plane of reference Π not going through C_1, so that we can write

$$\boldsymbol{\mathcal{P}}_1 \simeq \boldsymbol{\mathcal{P}}_0$$

and

$$\boldsymbol{\mathcal{P}}_i = \left[\mathbf{S}_{1i} + \mathbf{e}_{i1}\mathbf{r}_i^T \ \mu_i \mathbf{e}_{i1}\right] \quad i = 2, \ldots, n-1$$

As we have written it, each perspective projection matrix $\boldsymbol{\mathcal{P}}_i$, $i = 2, \ldots, n-1$ is defined by the special matrix $\mathbf{S}_{1i}$ and the epipole e_{i1} with view 1, the vector $\mathbf{r}_i$ representing the image of the line of intersection of the reference plane Π with the intrinsic plane of view i with respect to view 1 and the scalar μ_i. It will be sufficient for the rest of the discussion to assume that $\boldsymbol{\mathcal{P}}_i$ is given as

$$\boldsymbol{\mathcal{P}}_i = [\mathbf{H}_{1i,\Pi} \ \mu_i \mathbf{e}_{i1}]$$

where $\mathbf{H}_{1i,\Pi} = \mathbf{S}_{1i} + \mathbf{e}_{i1}\mathbf{r}_i^T$ represents the homography induced by Π between view 1 and i. Let us also note that, because of proposition 8.2, in the canonical frame of the first $n-1$ views, the plane Π is represented by the vector $\mathbf{\Pi} = [0,\ 0,\ 0,\ 1]^T$.

We now acquire an nth view. Given the three Fundamental matrices or the Trifocal tensor of two previous views i and j ($1 \leq i, j \leq n-1$) and the new view, we can construct a canonical representation for the triplet $(i,\ j,\ n)$.

In this representation, the perspective projection matrices $(\mathcal{P}'_i, \mathcal{P}'_j, \mathcal{P}'_n)$ are such that $\mathcal{P}'_i \simeq \mathcal{P}_0$ and $\mathcal{P}'_n$ and $\mathcal{P}'_j$ are given by the equations (11.2) and (11.3). In detail, we have:

$$\mathcal{P}'_j = \left[\mathbf{S}_{ij} + \mathbf{e}_{ji}\mathbf{r}'^T_j \; \mu'_j \mathbf{e}_{ji}\right]$$

and

$$\mathcal{P}'_n = \left[\mathbf{S}_{in} + \mathbf{e}_{ni}\mathbf{r}'^T_n \; \mu'_n \mathbf{e}_{ni}\right]$$

This is shown in figure 11.2. The vectors $\mathbf{r}'_j$ and $\mathbf{r}'_n$ on one hand, the two

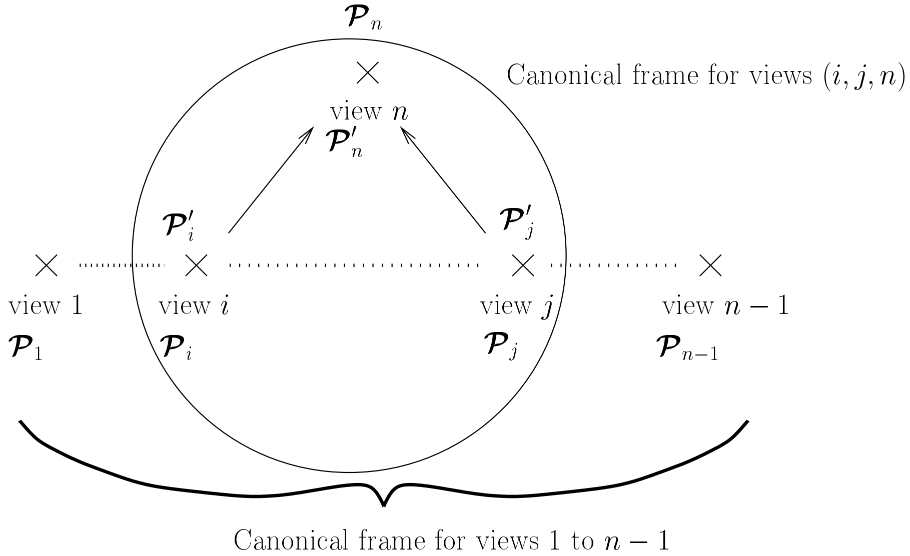


Figure 11.2: Stitching view n to views 1 to $n-1$.

scalars μ'_j and μ'_n on the other, are related as explained in propositions 11.4 and 11.5.

The problem is that the canonical representation of the first $n-1$ views provides two different expressions for the perspective projection matrices of the views i and j since the reference view is view 1, corresponding to two different projective bases. In order to "integrate" view n among the first $n-1$ views we need to choose the reference plane Π' that defines $\mathbf{r}'_j$ and $\mathbf{r}'_n$ to be the same, i.e. Π, as the one defined by $\mathbf{r}_i$ and $\mathbf{r}_j$. The difficulty is that $\mathbf{r}_j$ and $\mathbf{r}'_j$ are defined with respect to the first and the nth cameras, respectively, i.e. with respect to two different projective bases. Therefore, given $\mathbf{r}_j$, it is not completely obvious how to choose $\mathbf{r}'_j$. Note that since Π' does not go through C_i by definition, the same must be true of Π. This can be easily checked by considering the rank of $\mathbf{H}_{1i,\Pi}$. If necesssary, the reference plane should be changed and the canonical representation of the first $n-1$ views upgraded accordingly.

We denote by $H_{ij,\Pi'}$ the homography induced between the image planes i and j by the plane Π' defined by the vector $\mathbf{r}'_j$ or the vector $\mathbf{r}'_n$.

We show how to compute $\mathbf{r}'_j$ and μ'_j, i.e. how to choose the canonical frame of the triplet (i, j, n), so that when we return to the canonical frame of the first $n-1$ views we obtain the correct perspective projection matrices for views i and j. Let $\mathbf{H}_i$ be a matrix representing the change of projective coordinates from the canonical frame of the triplet (i, j, n) to the canonical frame of the first $n-1$ views. Such a matrix is given by:

$$\mathbf{H}_i = \begin{bmatrix} \boldsymbol{\mathcal{P}}_i \\ \boldsymbol{\Pi}^T \end{bmatrix}$$

We verify that $\boldsymbol{\mathcal{P}}_i \simeq \boldsymbol{\mathcal{P}}'_i \mathbf{H}_i \simeq \boldsymbol{\mathcal{P}}_0 \mathbf{H}_i$ and the condition we seek is

$$\boldsymbol{\mathcal{P}}_j \simeq \boldsymbol{\mathcal{P}}'_j \mathbf{H}_i$$

The main result is summarized in the following proposition

Proposition 11.6 *The nth view can be stitched to the first $n-1$ views linearly from the canonical representation of any three views (i, j, n), $i, j = 1, \ldots, n-1$. This means that the perspective projection matrix $\boldsymbol{\mathcal{P}}_n$ of the nth view in the canonical frame of the first $n-1$ views can be computed linearly from the canonical representations of the first $n-1$ views and of the views (i, j, n).*

Proof : Remember that, from proposition 6.16, the vector $\mathbf{r}'_j$ is the image in view i of the line of intersection of the intrinsic plane of camera j and the plane Π. Therefore we have:

$$\mathbf{r}'_j \simeq \tilde{\boldsymbol{\mathcal{P}}}_i(\boldsymbol{\Pi} \bigtriangleup \boldsymbol{\mathcal{P}}_j^T \mathbf{e}_{ji})$$

In detail, let $\boldsymbol{U}_i$, $\boldsymbol{V}_i$, $\boldsymbol{W}_i$ be the three row vectors of matrix $\boldsymbol{\mathcal{P}}_i$ and let us note $\boldsymbol{\Pi}_{ji}$ the vector $\boldsymbol{\mathcal{P}}_j^T \mathbf{e}_{ji}$. Because of proposition 5.11, we have:

$$\mathbf{r}'_j \simeq [\mid \boldsymbol{V}_i, \boldsymbol{W}_i, \boldsymbol{\Pi}, \boldsymbol{\Pi}_{ji} \mid, \mid \boldsymbol{W}_i, \boldsymbol{U}_i, \boldsymbol{\Pi}, \boldsymbol{\Pi}_{ji} \mid, \mid \boldsymbol{U}_i, \boldsymbol{V}_i, \boldsymbol{\Pi}, \boldsymbol{\Pi}_{ji} \mid]$$

Since all quantities on the right-hand side are known, this yields the direction of $\mathbf{r}'_j$, a unit vector that we note $\mathbf{r}'_{jN}$. We have therefore $\mathbf{r}'_j = \alpha'_j \mathbf{r}'_{jN}$. The scalar α'_j is computed by enforcing the following:

$$(\mathbf{S}_{ij} + \alpha'_j \mathbf{e}_{ji} \mathbf{r}'^T_{jN}) \mathbf{H}_{1i,\Pi} \simeq \mathbf{H}_{1j,\Pi} \tag{11.10}$$

which ensures that the diagram of figure 11.3 is commutative and yields α'_j linearly.

Note that the assumption that Π does not go through C_i is essential here because, were it not the case, the determinant of $\mathbf{H}_{1i,\Pi}$ would be 0 and, if the determinant of $\mathbf{H}_{1j,\Pi}$ was nonzero equation (11.10) would be impossible.

We have therefore determined $\mathbf{r}'_j$ and, because of proposition 11.4 or 11.5, this determines also $\mathbf{r}'_n$. Since, because of the same propositions, the ratio $\mu'_n : \mu'_j$ is also known, there remains to fix one of the ratios $\mu'_j : \mu_k$, $k = 2, \cdots, n-1$. We compute the ratio $\mu'_j : \mu_i$.

This is achieved by going back to the canonical system of the first $n-1$ views and multiplying $\boldsymbol{\mathcal{P}}'_i$, $\boldsymbol{\mathcal{P}}'_j$ (and of course $\boldsymbol{\mathcal{P}}'_n$) on the right by the 4×4 matrix

$$\mathbf{H}_i = \begin{bmatrix} \boldsymbol{\mathcal{P}}_i \\ \mathbf{\Pi}^T \end{bmatrix},$$

to obtain the original matrices $\boldsymbol{\mathcal{P}}_i$ and $\boldsymbol{\mathcal{P}}_j$. This is obviously true for view i. Taking into account the fact that $\mathbf{\Pi} = [0,\, 0,\, 0,\, 1]^T$, we find

$$\boldsymbol{\mathcal{P}}'_j \mathbf{H}_i = [(\mathbf{S}_{ij} + \mathbf{e}_{ji}\mathbf{r}'^T_j)\mathbf{H}_{1i,\Pi},\; \mu_i(\mathbf{S}_{ij} + \mathbf{e}_{ji}\mathbf{r}'^T_j)\mathbf{e}_{i1} + \mu'_j \mathbf{e}_{ji}]$$

Our choice of $\mathbf{r}'_j$ guarantees the existence of an α (and we can compute it

$$\begin{tikzcd}
1 \arrow[rr, "H_{1j,\Pi}"] \arrow[dr, "H_{1i,\Pi}"'] & & j \\
& i \arrow[ur, "H_{ij,\Pi'}"'] &
\end{tikzcd}$$

Figure 11.3: We choose $\mathbf{r}'_j$ in such a way that $H_{ij,\Pi'} \circ H_{1i,\Pi} = H_{1j,\Pi}$.

from equation (11.10)) such that

$$(\mathbf{S}_{ij} + \mathbf{e}_{ji}\mathbf{r}'^T_j)\mathbf{H}_{1i,\Pi} = \alpha \mathbf{H}_{1j,\Pi}$$

The ratio $\mu'_j : \mu_i$ is therefore obtained from the equation:

$$\mu_i(\mathbf{S}_{ij} + \mathbf{e}_{ji}\mathbf{r}'^T_j)\mathbf{e}_{i1} + \mu'_j \mathbf{e}_{ji} = \alpha \mu_j \mathbf{e}_{j1}$$

$\square$

11.3.3.2 The global method

The case of affine cameras What we call the global method is a generalization to the case of perspective projective cameras of a method developed for affine cameras by Tomasi and Kanade [365]. We first describe their method and then show how it can be extended to deal with the full perspective projection case. This extension is due to Triggs and Sturm [352].

We therefore consider n affine cameras (see chapter 5, section 5.4.2) and p 3D points M_j. Each perspective projection matrix $\boldsymbol{\mathcal{P}}_i$ takes the simpler form:

$$\boldsymbol{\mathcal{P}}_i = \begin{bmatrix} \mathbf{P}_i & \mathbf{p}_i \\ \mathbf{0}_3^T & 1 \end{bmatrix}$$

where $\mathbf{P}_i$ is a 2×3 matrix of rank 2 and $\mathbf{p}_i$ is a vector of size 2. Let $\mathbf{M}_j$ be the vector of size 3 representing the affine coordinates of the jth 3D point in some affine coordinate system and $\mathbf{m}_{ij}$ be the vector of size 2 representing the affine coordinates of its image in the ith view. The projection equation can be written:

$$\begin{bmatrix} \mathbf{m}_{ij} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{P}_i & \mathbf{p}_i \\ \mathbf{0}_3^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{M}_j \\ 1 \end{bmatrix} \tag{11.11}$$

We can further simplify the problem by taking the origin of the coordinates in each image plane at the center of gravity of the observed points. Because the cameras are affine, those centers of gravities are the images of the center of gravity of the 3D points which we can assume to be the origin of the 3D affine coordinate system (see chapter 5, section 5.4.2). This implies that the vectors $\mathbf{p}_i$ are all equal to zero.

The next thing we do is group all projection equations into a single large matrix equation:

$$\mathbf{W}_a = \begin{bmatrix} \mathbf{m}_{11} & \mathbf{m}_{12} & \dots & \mathbf{m}_{1p} \\ \mathbf{m}_{21} & \mathbf{m}_{22} & \dots & \mathbf{m}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{m}_{n1} & \mathbf{m}_{n2} & \dots & \mathbf{m}_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \\ \mathbf{P}_n \end{bmatrix} [\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_p] \tag{11.12}$$

The known $2n \times p$ matrix $\mathbf{W}_a$ appears as the product of a $2n \times 3$ matrix describing the unknown projection matrices and a $3 \times p$ matrix describing the unknown affine coordinates of the observed 3D points. Therefore, the rank of $\mathbf{W}_a$ is less than or equal to 3. Let $q = min(2n, p)$.

We compute the SVD of $\mathbf{W}_a$:

$$\mathbf{W}_a = \mathbf{U}_a \mathbf{\Sigma}_a \mathbf{V}_a^T$$

$\mathbf{U}_a$ is the $2n \times q$ matrix of the first q orthogonal eigenvectors of the symmetric matrix $\mathbf{W}_a \mathbf{W}_a^T$, $\mathbf{V}_a$ the $p \times q$ orthogonal matrix of the first q orthogonal eigenvectors of the symmetric matrix $\mathbf{W}_a^T \mathbf{W}_a$ and $\mathbf{\Sigma}_a$ is the $q \times q$ diagonal matrix whose diagonal elements are the singular values of the matrix $\mathbf{W}_a$. Since we have seen that $\mathbf{W}_a$ was of rank less than or equal to 3, there are only at most three nonzero such diagonal elements, noted σ_1^2, σ_2^2 and σ_3^2 which we can assume to be the first three elements on the diagonal of $\mathbf{\Sigma}_a$.

Because of this, only the first three columns of matrix $\mathbf{U}_a$ and the first three rows of $\mathbf{V}_a^T$ are relevant in the decomposition (we assume that $3 \leq q$. Let us note $\mathbf{U}'_a$ and $\mathbf{V}'^T_a$ the corresponding $2n \times 3$ and $3 \times p$ matrices. The SVD of $\mathbf{W}_a$ can be rewritten:

$$\mathbf{W}_a = \mathbf{U}'_a \mathbf{\Sigma}'_a \mathbf{V}'^T_a$$

where $\mathbf{\Sigma}'_a = diag(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. Let us write for example $\mathbf{U}''_a = \mathbf{U}'_a \mathbf{\Sigma}'_a$, the SVD of $\mathbf{W}_a$ is now:

$$\mathbf{W}_a = \mathbf{U}''_a \mathbf{V}'^T_a$$

The $2n \times 3$ matrix $\mathbf{U}''_a$ can be interpreted as describing a set of n affine projection matrices and the $3 \times p$ matrix $\mathbf{V}'^T_a$ as describing the affine coordinates of p 3D points.

Both the projection matrices and the 3D points are defined only up to an arbitrary affine transformation that leaves the origin of coordinates fixed. Indeed, such a transformation is defined by an invertible 3×3 matrix $\boldsymbol{\mathcal{A}}$ such that the points $\mathbf{M}_j$ are transformed to $\boldsymbol{\mathcal{A}}\mathbf{M}_j$ and the matrices $\mathbf{P}_j$ are transformed to $\mathbf{P}_j \boldsymbol{\mathcal{A}}^{-1}$. One verifies that $\mathbf{W}_a = (\mathbf{U}''_a \boldsymbol{\mathcal{A}}^{-1})(\boldsymbol{\mathcal{A}}\mathbf{V}'^T_a)$.

The reconstruction that has been obtained is defined up to an arbitrary affine transformation that preserves the origin of coordinates.

In practice, because of the uncertainty in the measurement of the coordinates of the points m_{ij}, the rank of $\mathbf{W}_a$ is greater than 3, in effect $\mathbf{W}_a$ is usually full rank with 3 singular values that are significantly larger than the others. If one sets to 0 all but those 3 singular values and computes the product $\mathbf{U}''_a \mathbf{V}'^T_a$, one obtains a matrix $\hat{\mathbf{W}}_a$ which is the best approximation of $\mathbf{W}_a$ by a matrix of rank 3 for the Frobenius norm [124], see figure 11.4. Let $\hat{\mathbf{P}}_i$ be the n projection matrices and $\hat{\mathbf{M}}_j$ be the p 3D points obtained from $\hat{\mathbf{W}}_a$. Since the Frobenius norm of the difference $\mathbf{W}_a - \hat{\mathbf{W}}_a$ is equal to $\sum_{i,j} d^2(\mathbf{m}_{ij}, \hat{\mathbf{P}}_i \hat{\mathbf{M}}_j)$ (equation (11.12)), we conclude that the choice of $\hat{\mathbf{W}}_a$ is optimal in the sense that it minimizes the reprojection error.

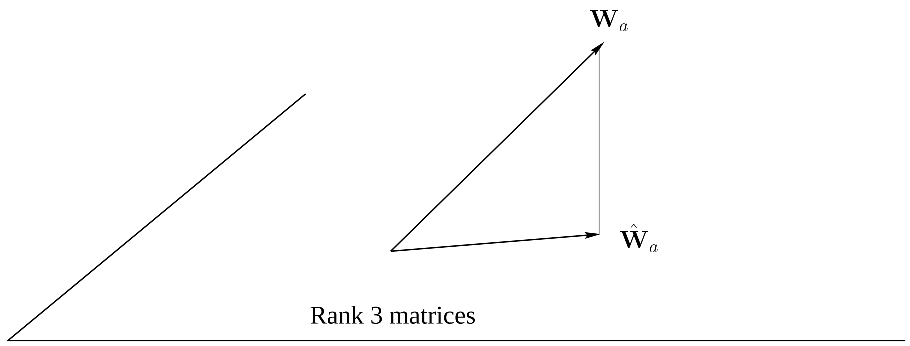


Figure 11.4: Among all rank 3 $2n \times p$ matrices $\mathbf{W}$, $\hat{\mathbf{W}}_a$ is the one that minimizes the Frobenius norm of the difference $\mathbf{W}_a - \mathbf{W}$.

Note that the method just described requires that all p points are visible in all n views to work. This assumption is often unrealistic in practice.

The case of perspective cameras In the case of perspective cameras, affine geometry does not suffice anymore: even if we choose a particular coordinate vector $\mathbf{m}_{ij}$ for each of the np observed points m_{ij}, a particular projective basis of 3D space, we still have to cope with the problem that the perspective projection matrices $\boldsymbol{\mathcal{P}}_i$ can be multiplied independently by an arbitrary scale factor as is also the case for the coordinate vectors $\mathbf{M}_j$ of the 3D points M_j.

Therefore, if we assume that we have chosen in each image a particular projective basis, for example the one that defines the usual pixel coordinates, and for each pixel m_{ij}, a particular coordinate vector, for example the one with the last coordinate equal to 1 as in equation (11.11), we must introduce some extra degrees of freedom to make the projective equality $\mathbf{m}_{ij} \simeq \boldsymbol{\mathcal{P}}_i \mathbf{M}_j$ an equality. We denote by λ_{ij} the np scale factors such that:

$$\lambda_{ij} \mathbf{m}_{ij} = \boldsymbol{\mathcal{P}}_i \mathbf{M}_j \tag{11.13}$$

The reason why we introduce those scale factors should be pretty obvious from the previous paragraph. In effect, they will allow us to write a projective analog of equation (11.12):

$$\mathbf{W}_p = \begin{bmatrix} \lambda_{11}\mathbf{m}_{11} & \lambda_{12}\mathbf{m}_{12} & \ldots & \lambda_{1p}\mathbf{m}_{1p} \\ \lambda_{21}\mathbf{m}_{21} & \lambda_{22}\mathbf{m}_{22} & \ldots & \lambda_{2p}\mathbf{m}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{n1}\mathbf{m}_{n1} & \lambda_{n2}\mathbf{m}_{n2} & \ldots & \lambda_{np}\mathbf{m}_{np} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mathcal{P}}_1 \\ \boldsymbol{\mathcal{P}}_2 \\ \vdots \\ \boldsymbol{\mathcal{P}}_n \end{bmatrix} [\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_p] \tag{11.14}$$

with the appropriate changes in the dimensions of the matrices: matrix $\mathbf{W}_p$ is now a $3n \times p$ matrix, and the two matrices in the right-hand side are of dimensions $3n \times 4$ and $4 \times p$.

Let us assume for a moment that we have been able to compute a set of λ_{ij}'s such that the equation (11.14) is true, the rest of the procedure that consists in computing a set of perspective projection matrices $\boldsymbol{\mathcal{P}}_i$, $i = 1, \ldots, n$ and of coordinate vectors $\mathbf{M}_j$, $j = 1, \ldots, p$, is very similar to the case of affine projection matrices described in the previous paragraph: because of the decomposition of $\mathbf{W}_p$ as the product of a $3n \times 4$ matrix and a $4 \times p$ matrix, its rank has to be less than or equal to 4; we compute its SVD:

$$\mathbf{W}_p = \mathbf{U}_p \boldsymbol{\Sigma}_p \mathbf{V}_p^T$$

Like in the affine case, $\mathbf{U}_p$ is the $2n \times q$ matrix of the orthogonal first q eigenvectors of the symmetric matrix $\mathbf{W}_p \mathbf{W}_p^T$, $\mathbf{V}_p$ the $p \times q$ orthogonal matrix of the orthogonal first q eigenvectors of the symmetric matrix $\mathbf{W}_p^T \mathbf{W}_p$ and $\boldsymbol{\Sigma}_p$ is the $q \times q$ diagonal matrix whose diagonal elements are the singular values of the matrix $\mathbf{W}_p$. Since we have seen that $\mathbf{W}_p$ was of rank less than or equal to 4, there are only at most four nonzero such diagonal elements, noted σ_1^2, σ_2^2, σ_3^2 and σ_4^2 which we can assume to be the first three elements on the diagonal of $\boldsymbol{\Sigma}_p$.

We then introduce, as in the affine case, the $2n \times 4$ matrix $\mathbf{U}'_p$ and the $4 \times p$ matrix $\mathbf{V}_p'^T$ which are the first four columns and the first four rows, respectively, of the matrices $\mathbf{U}_p$ and $\mathbf{V}_p^T$. The SVD of $\mathbf{W}_p$ can be rewritten:

$$\mathbf{W}_p = \mathbf{U}'_p \boldsymbol{\Sigma}'_p \mathbf{V}_p'^T$$

where $\boldsymbol{\Sigma}'_p = diag(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2)$. Let us write for example $\mathbf{U}''_p = \mathbf{U}'_p \boldsymbol{\Sigma}'_p$, the SVD of $\mathbf{W}_p$ is now:

$$\mathbf{W}_p = \mathbf{U}''_p \mathbf{V}_p'^T$$

The $2n \times 4$ matrix $\mathbf{U}''_p$ can be interpreted as describing a set of n perspective projection matrices and the $4 \times p$ matrix $\mathbf{V}_p'^T$ as describing the projective coordinates of p 3D points.

Both the projection matrices and the 3D points are defined only up to an arbitrary projective transformation. Indeed, such a transformation is defined by an invertible 4×4 matrix $\mathbf{H}$ such that the points $\mathbf{M}_j$ are transformed to $\mathbf{HM}_j$ and the matrices $\boldsymbol{\mathcal{P}}_j$ are transformed to $\boldsymbol{\mathcal{P}}_j\mathbf{H}^{-1}$. One verifies that $\mathbf{W}_p = (\mathbf{U}_p''\mathbf{H}^{-1})(\mathbf{H}\mathbf{V}_p'^T)$.

The reconstruction that has been obtained is defined up to an arbitrary projective transformation.

The difficulty in order to make this work is to choose the scale factors λ_{ij}. It turns out to be itself a difficult problem. Let us see why. It should be clear that equation (11.13) implies that:

$$\lambda_{ij}\mathbf{F}_{ik}\mathbf{m}_{ij} = \lambda_{kj}\mathbf{e}_{ki} \times \mathbf{m}_{kj} \tag{11.15}$$

for all pairs of views (i, k) and for all points M_j as soon as we compute $\mathbf{F}_{ik}$ and $\mathbf{e}_{ki}$ according to the formulas (6.8) and (6.9), respectively. This equation yields the ratio $\lambda_{ij} : \lambda_{kj}$. Assume now that we have estimated the $n-1$ Fundamental matrices $\mathbf{F}_{ii+1}$, $i = 1, \ldots, n-1$. Therefore, for a given 3D point, i.e. for a fixed value of j, we can compute linearly the ratios $\lambda_{1j} : \lambda_{2j} : \ldots : \lambda_{nj}$.

There are several problems with this simple approach. The first problem is that we have assumed that we had estimated the minimum number of Fundamental matrices, i.e. $n-1$, to allow us to compute $\lambda_{1j} : \lambda_{2j} : \ldots : \lambda_{nj}$. But what happens if we have estimated more, perhaps even all of them, i.e. $\binom{n}{2}$? It is unclear that the corresponding system of linear equations in $\lambda_{1j}, \lambda_{2j}, \ldots, \lambda_{nj}$ will have a solution!

Let us take as a toy example the case $n = 3$. We know from chapter 9 that the three Fundamental matrices $\mathbf{F}_{12}$, $\mathbf{F}_{23}$ and $\mathbf{F}_{31}$ must satisfy the three algebraic relations (9.3). If they do not, i.e. if our estimation procedure has not taken them into account, then the corresponding system of linear equations in $\lambda_{1j}, \lambda_{2j}, \ldots, \lambda_{nj}$ may not have a solution. In this case, we know how to avoid the problem, for example by computing the Fundamental matrix from one Trifocal tensor (see chapter 9) but for larger values of n we do not know at the time of this writing of any solution.

Another closely related problem is that we have no way of computing the Fundamental matrices and the epipoles according to equation (6.8) and (6.9) since this requires the knowledge of the perspective projection matrices which are one the things we are trying to compute. This means that in equation (11.15) there will be unwanted and unknown scale factors on the Fundamental matrix and the epipole. This is fine as long as we use only $n-1$ such equations because the scale factors are absorbed in the λ's. Problems arise if we use more equations. Returning to the previous example with $n = 3$, we see that even if the three Fundamental matrices and the six epipoles do satisfy the three relations (9.3), we still have to adjust the scale factors on those quantities in such a way that the six equations (9.4) and (9.5) are satisfied. That problem is in itself difficult and becomes worse when we increase the number of views and of Fundamental matrices.

Note in passing that in the case $n = 3$, the results given in proposition 11.5 allow us to solve the problem: given an estimated Trifocal tensor, recover Fundamental matrices and epipoles that satisfy all nine equations (9.3), (9.4) and (9.5) since we can compute three perspective projection matrices.

11.3.3.3 The bundle adjustment

We now describe a numerical method that can be applied in several contexts depending upon what information, projective, affine, or Euclidean is known. We limit ourselves to the computation of the projective canonical representation of n views and to the case where the observed primitives are points: we observe p points $M_1, \ldots, M_p$. For simplicity in the notations we assume that all p points are visible in all views albeit this is not necessary for the method to work. The image of the jth point in the ith view is noted m_{ij}.

The problem that we want to solve is the following. The methods described in the previous two sections provide us with estimates $(\boldsymbol{\mathcal{P}}_0, \hat{\boldsymbol{\mathcal{P}}}_1, \cdots, \hat{\boldsymbol{\mathcal{P}}}_n)$ of the perspective projection matrices and estimates $\hat{M}_j, j = 1, \cdots, p$ of the p 3-D points. In general the reprojected points $\hat{m}_{ij}, i = 1, \cdots, n, j = 1, \cdots, p$ represented by $\hat{\boldsymbol{\mathcal{P}}}_i \hat{\mathbf{M}}_j$ fail to be identical to the measurements m_{ij}. The bundle adjustment is a numerical method for adjusting the projection matrices and the 3-D points *starting from the initial estimate* provided by the results of sections 11.3.3.1 or 11.3.3.2 in such a way that the reprojected points become closer to the measured points.

Hence we consider the squared image distance $\varepsilon_{ij} = \frac{1}{2} d^2(m_{ij}, \hat{m}_{ij})$ between the measured pixel m_{ij} and the reprojected pixel $\hat{m}_{ij}$ and attempt to minimize simultaneously all the error measures ε_{ij} by adjusting the 3D points M_j and the parameters of the perspective projection matrices $\boldsymbol{\mathcal{P}}_i$.

If we stack the np numbers ε_{ij} into a large vector noted $\boldsymbol{\varepsilon}$, this vector is a function of the p points M_j and the parameters of the $n-1$ perspective projection matrices, i.e. a function of $N = 3p + 11(n-1)$ parameters.

The vector $\boldsymbol{\varepsilon}$ defines a function $\mathbf{f} : \mathbb{R}^N \longrightarrow \mathbb{R}^{np}$. A point in the space $\mathbb{R}^N$ represents p 3D points and $n-1$ perspective projection matrices. We have in general estimates $\hat{\boldsymbol{\mathcal{P}}}_i$, $i = 2, \cdots, n$ of the $n-1$ perspective projection matrices, built for example as explained in section 11.3.3.1 or 11.3.3.2. From those estimates and the pixels m_{ij} we can compute estimates $\hat{M}_j$ of the 3D points. All those define a vector $\hat{\mathbf{x}}$ in $\mathbb{R}^N$ such that $\mathbf{f}(\hat{\mathbf{x}}) = \hat{\boldsymbol{\varepsilon}}$. The central idea of the method is that it tries to change the vector $\hat{\mathbf{x}}$, i.e. to move the points $\hat{M}_j$ and to change the perspective projection matrices $\hat{\boldsymbol{\mathcal{P}}}_i$ in such a way as to minimize the Euclidean norm of the resulting vector $\boldsymbol{\varepsilon}$. In theory this norm should be zero, in practice only a local minimum for which not all coordinates of $\boldsymbol{\varepsilon}$ are equal to zero can be reached.

Technically, this is achieved as follows. Let $\mathbf{x}_0 = \hat{\mathbf{x}}$ and $\boldsymbol{\varepsilon}_0 = \mathbf{f}(\mathbf{x}_0)$. At iteration k, $k = 1, \cdots$, we consider $\mathbf{f}(\mathbf{x}_{k-1} + \boldsymbol{\Delta})$ where the increment vector $\boldsymbol{\Delta}$ is small and we expand $\mathbf{f}$ up to first order at the point $\mathbf{x}_{k-1}$:

$$\mathbf{f}(\mathbf{x}_{k-1} + \boldsymbol{\Delta}) = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{J}_{k-1}\boldsymbol{\Delta} + O(\|\boldsymbol{\Delta}\|)$$

$\mathbf{J}_{k-1}$ is the Jacobian matrix, i.e. the first-order derivative, of the function $\mathbf{f}$ evaluated at the point $\mathbf{x}_{k-1}$. Its exact form depends upon the parameterization of the perspective projection matrices $\boldsymbol{\mathcal{P}}_i$ but it has a very particular structure that we will describe shortly. But before we do this, let us proceed with the determination of the increment vector $\boldsymbol{\Delta}$.

$\boldsymbol{\Delta}$ is chosen so as to minimize the norm of $\mathbf{f}(\mathbf{x}_{k-1} + \boldsymbol{\Delta})$ up to first order, i.e. it is chosen so that it minimizes the norm of $\mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{J}_{k-1}\boldsymbol{\Delta}$. This is a standard linear least-squares problem whose solution is obtained by solving the linear system of equations:

$$\mathbf{J}_{k-1}^T \mathbf{J}_{k-1}\boldsymbol{\Delta} + \mathbf{J}_{k-1}^T \boldsymbol{\varepsilon}_{k-1} = \mathbf{0} \tag{11.16}$$

These equations are called the *normal equations*. We have written $\boldsymbol{\varepsilon}_{k-1}$ for $\mathbf{f}(\mathbf{x}_{k-1})$. Let $\boldsymbol{\Delta}_{k-1}$ be a solution of that system, we then define

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \boldsymbol{\Delta}_{k-1}$$

and iterate the procedure until convergence is reached.

In practice several problems can occur such as convergence to a local minimum where $\boldsymbol{\varepsilon}$ is not equal to $\mathbf{0}$, or even divergence if the first-order approximation is not good enough. It has been reported many times in the literature that the results are very sensitive to the initial point $\mathbf{x}_0$ and that it is very important that this point is close to a solution. Therefore the computation of an initial "good" estimate is essential.

Let us now study in slightly more detail the structure of the Jacobian matrix $\mathbf{J}$. It turns out that this matrix is very sparse. This can be seen as follows. Each coordinate of $\mathbf{f}(\mathbf{x})$ is one of the $\boldsymbol{\varepsilon}_{ij}$ and depends only upon he matrix $\boldsymbol{\mathcal{P}}_i, i = 2, \cdots, n$ and the 3D point M_j. Each row of $\mathbf{J}$ contains the partial derivatives of the corresponding ε_{ij} with respect to all variables, the perspective projection matrices and the 3D points. But because of the previous remark, only the partial derivatives of ε_{ij} with respect to the parameters of $\boldsymbol{\mathcal{P}}_i$ and the coordinates of M_j will be non zero. We show as an example the case where we have three views ($n = 3$) of four points ($p = 4$) in figure 11.5.

outbox

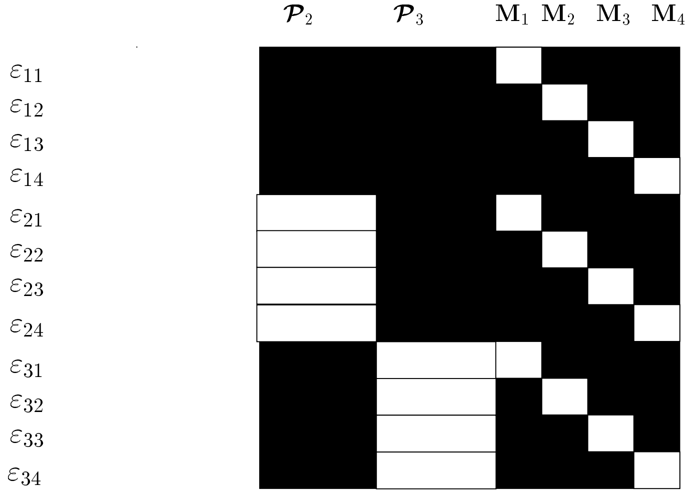


Figure 11.5: The $np \times N$ Jacobian matrix $\mathbf{J}$ is sparse. The nonzero parts are represented in white, the zero parts in black. For example, since ε_{11} is solely a function of M_1, only its partial derivatives with respect to the coordinates of M_1 are nonzero.

Having done this, we can make more precise the structure of the $N \times N$ matrix $\mathbf{J}^T\mathbf{J}$. We note $\dfrac{\partial \varepsilon_{ij}}{\partial \boldsymbol{\mathcal{P}}_i}$ the partial derivative of ε_{ij} with respect to the parameters defining the projection matrix $\boldsymbol{\mathcal{P}}_i$. This is a 1×11 matrix in the case where we parameterize $\boldsymbol{\mathcal{P}}_i$ by its elements. Similarly, we note $\dfrac{\partial \varepsilon_{ij}}{\partial \mathbf{M}_j}$ the partial derivative of ε_{ij} with respect to the coordinates of the 3D point $\mathbf{M}_j$. It is a 1×3 matrix. The structure of the symmetric matrix $\mathbf{J}^T\mathbf{J}$ is most easily described by introducing the following sets of matrices:

- The $n-1$ 11×11 symmetric matrices $\mathbf{U}_i$ defined by:

$$\mathbf{U}_i = \sum_{j=1}^{p} (\frac{\partial \hat{\mathbf{m}}_{ij}}{\partial \boldsymbol{\mathcal{P}}_i})^T \mathbf{C}_{ij} (\frac{\partial \hat{\mathbf{m}}_{ij}}{\partial \boldsymbol{\mathcal{P}}_i}) \quad i = 2, \cdots, n$$

- The p 3×3 symmetric matrices $\mathbf{V}_j$ defined by:

$$\mathbf{V}_j = \sum_{i=1}^{n} (\frac{\partial \hat{\mathbf{m}}_{ij}}{\partial \mathbf{M}_j})^T \mathbf{C}_{ij} (\frac{\partial \hat{\mathbf{m}}_{ij}}{\partial \mathbf{M}_j})$$

- The $(n-1)p$ 11×3 matrices $\mathbf{W}_{ij}$ defined by:

$$\mathbf{W}_{ij} = (\frac{\partial \hat{\mathbf{m}}_{ij}}{\partial \boldsymbol{\mathcal{P}}_i})^T \mathbf{C}_{ij} (\frac{\partial \hat{\mathbf{m}}_{ij}}{\partial \mathbf{M}_j}) \quad i = 2, \cdots, n, \; j = 1, \cdots, p$$

The matrix noted $\mathbf{C}_{ij}$ above is the following 2×2 matrix:

$$\mathbf{C}_{ij} = (\hat{\mathbf{m}}_{ij} - \mathbf{m}_{ij})(\hat{\mathbf{m}}_{ij} - \mathbf{m}_{ij})^T$$

The matrix $\mathbf{J}^T\mathbf{J}$ is built from those three classes of matrices as indicated in figure 11.6.

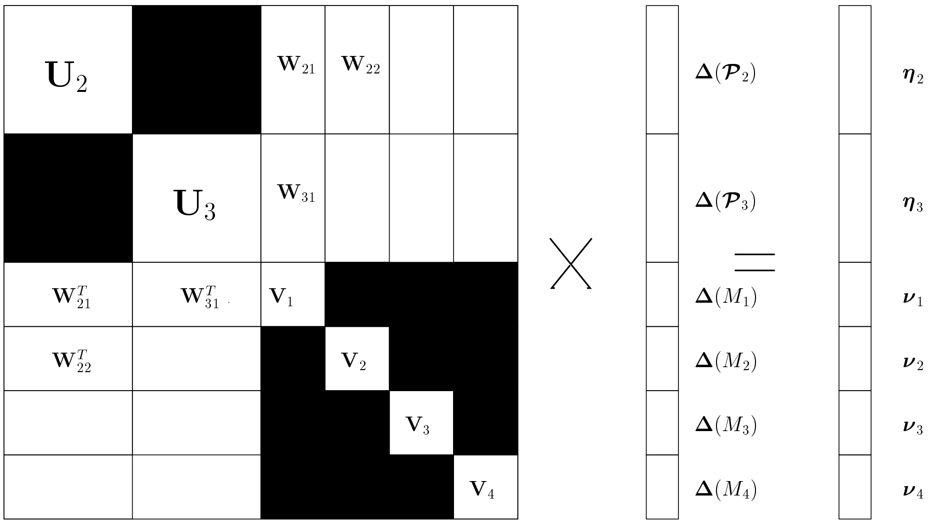


Figure 11.6: Normal equations.

In order to match the structure of $\mathbf{J}^T\boldsymbol{\varepsilon}_{k-1}$ to that of $\mathbf{J}^T\mathbf{J}$ it is useful to introduce the following two sets of matrices:

- The n vectors $\boldsymbol{\eta}_i$ of size 11 defined by:

$$\boldsymbol{\eta}_i = \sum_{j=1}^{p} (\frac{\partial \hat{\mathbf{m}}_{ij}}{\partial \boldsymbol{\mathcal{P}}_i})^T (\mathbf{m}_{ij} - \hat{\mathbf{m}}_{ij}) \varepsilon_{ij} \quad i = 2, \cdots, n$$

- The p vectors $\boldsymbol{\nu}_j$ of size 3 defined by:

$$\boldsymbol{\nu}_j = \sum_{i=1}^{n} (\frac{\partial \hat{\mathbf{m}}_{ij}}{\partial \mathbf{M}_j})^T (\mathbf{m}_{ij} - \hat{\mathbf{m}}_{ij}) \varepsilon_{ij} \quad j = 1, \cdots, p$$

The normal equations (11.16) can now be rewritten in a form that clearly shows the increments relative to the changes in the perspective projection matrices, noted $\boldsymbol{\Delta}(\boldsymbol{\mathcal{P}})$, and to the changes in the 3D points, noted $\boldsymbol{\Delta}(\mathbf{M})$:

$$\begin{bmatrix} \mathbf{U} & \mathbf{W} \\ \mathbf{W}^T & \mathbf{V} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Delta}(\boldsymbol{\mathcal{P}}) \\ \boldsymbol{\Delta}(\mathbf{M}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\eta}(\boldsymbol{\mathcal{P}}) \\ \boldsymbol{\nu}(\mathbf{M}) \end{bmatrix} \tag{11.17}$$

Assume that the $3p \times 3p$ block-diagonal matrix $\mathbf{V}$ is invertible and multiply both sides of (11.17) on the left by the matrix

$$\begin{bmatrix} \mathbf{I}_{11(n-1)} & -\mathbf{W}\mathbf{V}^{-1} \\ \mathbf{0}_{3p \times 11(n-1)} & \mathbf{I}_{3p} \end{bmatrix}$$

we obtain

$$\begin{bmatrix} \mathbf{U} - \mathbf{W}\mathbf{V}^{-1}\mathbf{W}^T & \mathbf{0}_{3p} \\ \mathbf{W}^T & \mathbf{V} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta}(\boldsymbol{\mathcal{P}}) \\ \boldsymbol{\Delta}(\mathbf{M}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\eta}(\boldsymbol{\mathcal{P}}) - \mathbf{W}\mathbf{V}^{-1}\boldsymbol{\nu}(\mathbf{M}) \\ \boldsymbol{\nu}(\mathbf{M}) \end{bmatrix}$$

This separates the block of variables $\boldsymbol{\Delta}(\boldsymbol{\mathcal{P}})$ from $\boldsymbol{\Delta}(\mathbf{M})$:

$$\begin{aligned} (\mathbf{U} - \mathbf{W}\mathbf{V}^{-1}\mathbf{W}^T)\boldsymbol{\Delta}(\boldsymbol{\mathcal{P}}) &= \boldsymbol{\eta}(\boldsymbol{\mathcal{P}}) - \mathbf{W}\mathbf{V}^{-1}\boldsymbol{\nu}(\mathbf{M}) \\ \boldsymbol{\Delta}(\mathbf{M}) &= \mathbf{V}^{-1}(\boldsymbol{\nu}(\mathbf{M}) - \mathbf{W}^T\boldsymbol{\Delta}(\boldsymbol{\mathcal{P}}) \end{aligned}$$

The first equation yields $\boldsymbol{\Delta}(\boldsymbol{\mathcal{P}})$, the second $\boldsymbol{\Delta}(\mathbf{M})$.

11.3.3.4 An example of the recovery of the projective stratum

The 11 images of figure 11.7 are part of a sequence of 45 images that have acquired with a Kodak digital camera DCS160 at various points of the INRIA campus in Sophia-Antipolis. The Fundamental matrices have been estimated between as many pairs of views as possibleusing the methods described in chapter 7. The results for a subset of six images are shown in table 11.1.

This table shows in its upper-diagonal part the residuals, in pixels, i.e. the square root of the value of the expression 7.14 divided by the number of pixels used in the estimation of the Fundamental matrix which is shown in its lower-diagonal part.

Similar results are shown in table 11.2 for the six views of a different scene shown in figure 11.8.

	i.3	i.4	i.5	i.6	i.7	i.8
i.3		0.338	0.442	0.242	0.357	0.469
i.4	27		0.31	0.209	0.376	0.553
i.5	27	73		0.248	0.264	0.635
i.6	36	22	35		0.209	0.155
i.7	29	42	50	41		0.291
i.8	14	33	34	26	57	

Table 11.1: The Fundamental matrixes between all $\binom{6}{2}$ pairs of views among the images 3 to 8 of figure 11.7. The upper diagonal part of the matrix shows the residuals in pixels. The lower diagonal part shows the number of pairs of matched points that have been used in the estimation of the Fundamental matrixes: Courtesy **Sylvain Bougnoux**, INRIA.

	i.0	i.1	i.2	i.3	i.4	i.5
i.0		0.91	0.73	0.59	0.74	0.62
i.1	40		0.45	0.4	0.5	0.67
i.2	41	70		0.36	0.45	0.38
i.3	40	66	71		0.41	0.31
i.4	40	67	72	71		0.34
i.5	45	61	66	64	81	

Table 11.2: The Fundamental matrixes between all $\binom{6}{2}$ pairs of views among the images of figure 11.8. The upper diagonal part of the matrix shows the residuals in pixels. The lower diagonal part shows the number of pairs of matched points that have been used in the estimation of the Fundamental matrixes: Courtesy **Sylvain Bougnoux**, INRIA.

Figure 11.7: Eleven out of fourty five views of the INRIA site in Sophia-Antipolis: Courtesy **Sylvain Bougnoux**, INRIA.

Figure 11.8: Six out of ? views of the Arcades square in Valbonne: Courtesy **Sylvain Bougnoux**, INRIA.

11.4 Affine and Euclidean strata

We have seen in chapter 8, section 8.3.4, how to recover the infinity homography between two views. Proposition 8.9 gives us the form of the one-parameter family of affine canonical forms for the two perspective projection matrices. The analysis that has been done in this chapter, section 11.3, shows that the projective canonical form of an arbitrary number, n, of views can be turned into an affine canonical form as soon as the infinity homography is known between any pair of the n views. We assume for simplicity but without loss of generality that this homography is known between the views 1 and 2. We obtain the following proposition:

Proposition 11.7 *Given the projective canonical form* $(\boldsymbol{\mathcal{P}}_0, \boldsymbol{\mathcal{P}}_2, \ldots, \boldsymbol{\mathcal{P}}_n)$ *of n views and the infinity homography matrix* $\mathbf{H}_{\infty 12}$ *between the views 1 and 2, the affine canonical form of the n views is of the form:*

$$\boldsymbol{\mathcal{P}}_i^A = [\mathbf{H}_{\infty 1i}\, \mu_i \mathbf{e}_{i1}] \quad i = 2, \ldots, n$$

$\mathbf{H}_{\infty 1i}$ *is a representation of the infinity homography between views 1 and i* *which can be computed from* $\mathbf{H}_{\infty 12}$ *for* $i \geq 3$*:*

$$\mathbf{H}_{\infty 1i} = \mathbf{S}_{1i} + \mathbf{e}_{i1}\mathbf{r}_{\infty i}^T$$

μ_i *is a scalar that can be computed from* μ_2 *for* $i \geq 3$.

Proof : The proof is exactly the same as in propositions 11.4, 11.5 and 11.6. For example, in propositions 11.4 and 11.5, the vector $\mathbf{r}_2$ corresponds in our context to the plane at infinity Π_∞ and is obtained from definition 6.6, proposition 6.16 and the matrix $\mathbf{H}_{\infty 12}$ of the infinity homography between views 1 and 2. The propositions 11.4, 11.5 and 11.6 explain how to compute $(\mathbf{r}_{\infty 3}, \mu_3), \ldots, (\mathbf{r}_{\infty n}, \mu_n)$. □

We extend the previous result to the Euclidean setting as follows. We have seen in chapter 12 several methods to recover the intrinsic parameters of a camera. We will see more in sections 12.4 and 12.5. We now state and prove the general result that if we are given an affine canonical representation of n cameras, the knowledge of the intrinsic parameters of one of them is sufficient to obtain the Euclidean canonical representation of those n views.

Proposition 11.8 *Given the affine canonical form* $(\boldsymbol{\mathcal{P}}_0, \boldsymbol{\mathcal{P}}_2^A, \ldots, \boldsymbol{\mathcal{P}}_n^A)$ *of* n *views and the matrix of the intrinsic parameters* $\mathbf{A}_i$, $1 \leq i \leq n$ *of any of the cameras, the Euclidean canonical form of the* n *views is of the form:*

$$\boldsymbol{\mathcal{P}}_k = \begin{cases} \mathbf{A}_1\boldsymbol{\mathcal{P}}_0 & \textit{if} \quad k = 1 \\ \mathbf{A}_k\,[\mathbf{R}_k\; \mu_k\mathbf{t}_{k1}] & \textit{if} \quad k = 2, \ldots, n \end{cases}$$

$\mathbf{R}_k$ *is the matrix of the relative rotation between view 1 and view* k, $\mathbf{A}_k$ *is the matrix of the intrinsic parameters of camera* k. $\mathbf{t}_{k1}$ *is the vector* $\overrightarrow{C_kC_1}$ *expressed in the camera normalized coordinate system of the* k*th camera.*

Proof : We assume without loss of generality that $i = 1$. According to proposition 11.7, $\boldsymbol{\mathcal{P}}_k^A = [\mathbf{H}_{\infty 1k}\; \mu_k\mathbf{e}_{k1}]$. Let $\mathbf{K}$ be the Kruppa matrix of the kth view for $k = 1, \ldots, n$. Equation (12.1) of chapter 12 yields:

$$\mathbf{K}_k \simeq \mathbf{H}_{\infty 1k}\mathbf{K}_1\mathbf{H}_{\infty 1k}$$

and $\mathbf{A}_k$ is obtained by Cholesky decomposition of $\mathbf{K}_k$. $\mathbf{R}_k$ is then obtained from proposition 6.25 of chapter 6:

$$\mathbf{R}_k = \alpha_k\mathbf{A}_k^{-1}\mathbf{H}_{\infty 1k}\mathbf{A}_1$$

where $\alpha_k = det(\mathbf{A}_k)(det(\mathbf{A}_1\, det(\mathbf{H}_{\infty 1k}))^{-1}$.

Having computed the matrices $\mathbf{A}_k$ and $\mathbf{R}_k$, we can scale the perspective projection matrices $\boldsymbol{\mathcal{P}}_k^A$, $k = 2, \ldots, n$ so that they are equal to $\left[\mathbf{A}_k \mathbf{R}_k \mathbf{A}_1^{-1} \; \mu'_k \mathbf{e}'_{k1}\right]$. Let us further note that all the ratios $\mu'_2 : \ldots : \mu'_n$ are known (propositions 11.6 and 11.7). Therefore we have $\mathbf{e}'_{k1} = \mu \mathbf{A}_k \mathbf{t}_{k1}$. If we now apply the affine change of coordinate system defined by the matrix:

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{0}_3 \\ \mathbf{0}_3^T & 1 \end{bmatrix},$$

we obtain the Euclidean canonical form with $\mu_k = \mu'_k \mu$. $\square$

11.5 Stereo rigs

In this section, we consider a moving stereo rig. We study the case when the two cameras which compose the rig are rigidly attached, and their intrinsic parameters are constant. When doing a displacement of the rig, the four resulting views (the left and right view before and after the displacement) satisfy additional constraints, unlinke the four views obtained from a single moving camera which in general are unconstrained. This makes it possible to recover the affine stratum, as described in Sec. 11.5.1 and using techniques described in the next chapter, the Euclidean stratum.

Using a stereo rig instead of a single moving camera has also the advantage that because the epipolar geometry between the left and right camera is fixed, one can use all the image pairs taken in a sequence to estimate the Fundamental matrix, therefore accumulating point matches for a better estimation. This is especially interesting when the scene is close to a critical configuration, because with generic motion, repetitions of critical configurations are no longer critical.

On the other hand, in some cases it might be easier to obtain monocular correspondences by tracking points over time in the left camera and in the right camera. In this situation, one can use the redundancy to recover the epipolar geometry of the stereo rig without using correspondences between the left and right cameras. This is described in the Euclidean case in Sec. 11.5.2.

11.5.1 Affine calibration

We consider in this section a moving stereo rig and suppose that the same set of points are observed before and after the displacement. We first determine the epipolar geometry of the stereo rig between the left and right camera, and perform a projective reconstruction of this set of points before the displacement, let us say in $(\Pi, 1)$, a projective canonical frame of the stereo rig that we note $\mathcal{B}_1$ for simplicity. We do the same operation after the displacement, in the *same* projective canonical frame, noted $\mathcal{B}_2$ after the displacement. Since the ambiguity in reconstruction is at most projective, these two reconstructions are necessarily related by an homography $\boldsymbol{\mathcal{H}}$ of $\mathbb{P}^3$, which can be determined from the two sets of reconstructed points. In this section, we will see that we can extract the affine calibration from $\boldsymbol{\mathcal{H}}$, using a technique developped by [166] that we follow.

Let us now consider $\mathcal{E}_1$ and $\mathcal{E}_2$ a Euclidean coordinate system attached to the stereo rig before and after the displacement. Because we have used the same projective canonical frame before and after displacement, the form of the transformation from the projective coordinate system to the Euclidean coordinate system is known from (8.18):

$$\mathbf{Q}_{\mathcal{B}_1}^{\mathcal{E}_1} = \begin{bmatrix} \mathbf{A}^{-1} & 0 \\ -\mathbf{r}_\infty^T & 1 \end{bmatrix} \tag{11.18}$$

where $\mathbf{A}$ is the matrix of intrinsic parameters of the left camera, and $\mathbf{r}_\infty$ is the intrinsic representation of the plane at infinity as defined by Eq. (6.28). This expression depends only on the intrinsic and extrinsic parameters of the stereo rig. Since they are kept fixed, so is $\mathbf{Q}_{\mathcal{B}_1}^{\mathcal{E}_1} \equiv \mathbf{Q}_{\mathcal{B}_2}^{\mathcal{E}_2}$. Since $\mathcal{E}_1$ and $\mathcal{E}_2$ are Euclidean coordinate systems, the transformation between them is a displacement. Therefore $\boldsymbol{\mathcal{H}}$ is conjugate to the rigid displacement between the two cameras of the stereo rig, as illustrated by Fig. 11.9:

$$\boldsymbol{\mathcal{H}} = \mathbf{Q}_{\mathcal{E}_1}^{\mathcal{B}_1} \boldsymbol{\mathcal{D}} \mathbf{Q}_{\mathcal{B}_1}^{\mathcal{E}_1} \tag{11.19}$$

with

$$\boldsymbol{\mathcal{D}} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}$$

Note that we have used the symbol "equal" rather than "proportional" in this expression. This is because the scale factor can be recovered exacly by noting that two conjugate matrices have the same determinant and trace. The trace of a displacement is always positive, therefore the scale of $\boldsymbol{\mathcal{H}}$ is $\mathrm{sign}(\mathrm{trace}\boldsymbol{\mathcal{H}})(\det\boldsymbol{\mathcal{H}})^{1/4}$. After dividing $\boldsymbol{\mathcal{H}}$ by this value the previous equality is exact.

Since the affine calibration of a camera undergoing a pure translation has been already studied in Sec. 8.3.4.2, in this section we concentrate on displacements which have a non-null rotation. The following proposition makes it possible to extract the intrinsic representation of the plane at infinity $\mathbf{r}_\infty$ from a single general displacement of the stereo rig:

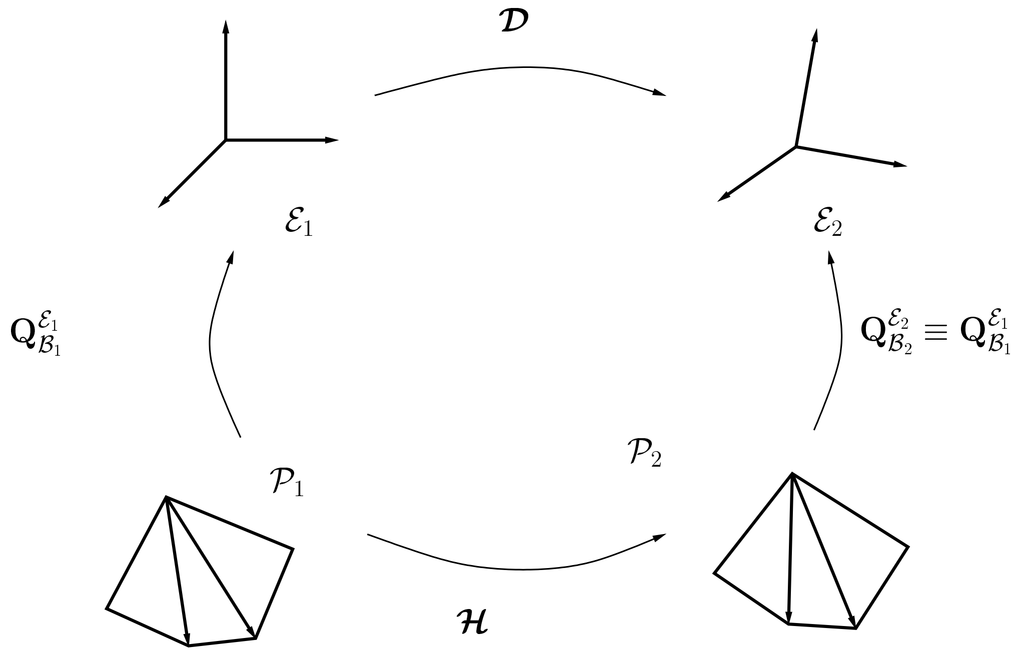


Figure 11.9: Motion of a stereo rig. The homography of $\mathbb{P}^3$ relating the projective reconstructions before and after the motion is conjugate to a rigid displacement: $\boldsymbol{\mathcal{H}} = \mathbf{Q}_{\mathcal{E}_1}^{\mathcal{B}_1} \boldsymbol{\mathcal{D}} \mathbf{Q}_{\mathcal{B}_1}^{\mathcal{E}_1}$.

Proposition 11.9 *Given a stereo rig of fixed geometry and intrinsic parameters, $\boldsymbol{\mathcal{H}}$ the homography of $\mathbb{P}^3$ between the projective representation in the canonical frame before and after a displacement which is not a pure translation, the vector $\mathbf{\Pi}_\infty = \begin{bmatrix} -\mathbf{r}_\infty^T \\ 1 \end{bmatrix}$ is the intrinsic representation of the plane at infinity for the pair of cameras in the stereo rig. It is an eigenvector of the matrix $\boldsymbol{\mathcal{H}}^{-T}$ associated with the eigenvalue 1. If the displacement is not a planar motion, this eigenvector is unique, otherwise there is a second independant eigenvector associated with the eigenvalue 1, $\begin{bmatrix} \mathbf{A}^{-T}\mathbf{r} \\ 0 \end{bmatrix}$, where $\mathbf{A}$ is the matrix of the intrinsic parameters of the left camera, and $\mathbf{r}$ is a unit vector parallel to the axis of the rotation of the displacement $\boldsymbol{\mathcal{D}}$ of the rig.*

Proof :

The eigenvalues associated with a rigid displacement are $\lambda \in \{e^{i\theta}, e^{-i\theta}, 1, 1\}$ where θ is the angle of rotation. The geometric multiplicity of an eigenvalue is equal to the dimension of its associated eigenspace. Let $\mathbf{r}$ be the rotation axis and $\mathbf{t}$ be the translation. The geometric multiplicity of $\lambda = 1$ depends on the motion type: it is 1 for general motion ($\mathbf{r}^T\mathbf{t} \neq 0$), and 2 for planar motion ($\mathbf{r}^T\mathbf{t} = 0$).

The eigenvalues of $\boldsymbol{\mathcal{D}}^{-T}$ are $1/\lambda$, and therefore among them is the value 1. Let's determine the eigenspace associated with this eigenvalue. Vectors $\mathbf{Y} = \begin{bmatrix} \mathbf{y} \\ \gamma \end{bmatrix}$ of this eigenspace satisfy $\boldsymbol{\mathcal{D}}^{-T}\mathbf{Y} = \mathbf{Y}$, which splits into the two equations:

$$\begin{aligned} \mathbf{R}\mathbf{y} &= \mathbf{y} \\ \mathbf{t}^T\mathbf{R}\mathbf{y} &= 0 \end{aligned}$$

The first equation has two solutions, $\mathbf{y} = \mathbf{r}$ or $\mathbf{y} = 0$. For a general motion, only the second solution is possible for the system, and therefore the eigenspace is the 1-dimensional space spanned by $\mathbf{Y}_1 = [0, 0, 0, 1]^T$, which represents the plane at infinity. For a planar motion, both solutions are possible for the system, and the eigenspace is the 2-dimensional space spanned by $\mathbf{Y}_1$ and $\mathbf{Y}_2 = \begin{bmatrix} \mathbf{r} \\ 0 \end{bmatrix}$, the plane orthogonal to the rotation axis and passing through the origin.

Transposing Eq. (11.19) yields:

$$\boldsymbol{\mathcal{H}}^{-T} = \mathbf{Q}_{\mathcal{B}_1}^{\mathcal{E}_1\,T} \boldsymbol{\mathcal{D}}^{-T} \mathbf{Q}_{\mathcal{B}_1}^{\mathcal{E}_1\,-T}$$

from which we obtain:

$$\boldsymbol{\mathcal{H}}^{-T}(\mathbf{Q}_{\mathcal{B}_1}^{\mathcal{E}_1\,T}\mathbf{Y}) = \mathbf{Q}_{\mathcal{B}_1}^{\mathcal{E}_1\,T}\boldsymbol{\mathcal{D}}^{-T}\mathbf{Y}$$

We conclude that the eigenvectors $\mathbf{X}$ of $\boldsymbol{\mathcal{H}}^{-T}$ are obtained from the eigenvectors $\mathbf{Y}$ of $\boldsymbol{\mathcal{D}}^{-T}$ by $\mathbf{X} = \mathbf{Q}_{\mathcal{B}_1}^{\mathcal{E}_1\,T}\mathbf{Y}$. Using the expression of $\mathbf{Q}_{\mathcal{B}_1}^{\mathcal{E}_1}$ in Eq. (11.18) gives the announced result. □

We have studied a single displacement of the stereo rig. We now extend our study to the case of multiple displacements. This has two interesting consequences. First, the advantage of using a stereo rig rather than a single camera is that not only it is possible to recover the plane at infinity, but also we can do so by combining any number of motions, the eigenvector being the common root of a set of linear equations, which in practice are solved by a linear least-squares procedure. Second, the previous proposition does not give us an affine calibration in the case of planar motion, however by combining the constraints derived from two such planar motions with distinct rotation axes, we obtain an unique solution.

Proposition 11.10 *If $\boldsymbol{\mathcal{H}}_1 \ldots \boldsymbol{\mathcal{H}}_n$ are homographies of $\mathbb{P}^3$ arising from motions of a fixed stereo rig which include either one general motion or two planar motions with different rotation axes, then the unique non trivial solution of the system:* $\begin{bmatrix} \boldsymbol{\mathcal{H}}_1^T - \mathbf{I}_4 \\ \vdots \\ \boldsymbol{\mathcal{H}}_n^T - \mathbf{I}_4 \end{bmatrix} \mathbf{X} = \mathbf{0}$ *is the vector* $\mathbf{\Pi}_\infty = \begin{bmatrix} -\mathbf{r}_\infty^T \\ 1 \end{bmatrix}$.

Proof : We notice that $\boldsymbol{\mathcal{H}}^T$ and $\boldsymbol{\mathcal{H}}^{-T}$ have the same eigenvector associated with $\lambda = 1$. The vector $\mathbf{r}_\infty$ depends only on the geometry of the stereo rig, not on its motion, therefore it satisfies all the equations of the system. The only thing left to prove is that if we have two planar motions with distinct rotation axis $\mathbf{r}_1$ and $\mathbf{r}_2$, the only common eigenvector for the eigenvalue 1 is $\mathbf{\Pi}_\infty$. Such a common eigenvector would satisfy:

$$\lambda_1 \begin{bmatrix} -\mathbf{r}_\infty^T \\ 1 \end{bmatrix} + \mu_1 \begin{bmatrix} \mathbf{A}^{-T}\mathbf{r}_1 \\ 0 \end{bmatrix} = \lambda_2 \begin{bmatrix} -\mathbf{r}_\infty^T \\ 1 \end{bmatrix} + \mu_2 \begin{bmatrix} \mathbf{A}^{-T}\mathbf{r}_2 \\ 0 \end{bmatrix}$$

After some simplifications, we find that this would imply $\mathbf{A}(\mu_1\mathbf{r}_1 + \mu_2\mathbf{r}_2) = 0$, and since $\mathbf{A}$ is invertible, $\mu_1\mathbf{r}_1 + \mu_2\mathbf{r}_2 = 0$. Since $\mathbf{r}_1$ and $\mathbf{r}_2$ are not proportional, $\mu_1 = \mu_2 = 0$, and therefore the common eigenvector has to be Π_s. □

We have therefore seen that from the transformations $\boldsymbol{\mathcal{H}}$ it is possible to extract the vector $\mathbf{r}_\infty$ which encodes the position of the plane at infinity in the chosen projective canonical frame of the two views of the stereo rig. This gives directly the infinity homography of the stereo rig from its Fundamental matrix using Eq. (6.28).

We can go even further if we consider the two positions of the first camera of the rig (resp. of the second camera of the rig). The plane at infinity induces an homography, noted $\mathbf{H}_{1\infty}$ (resp. $\mathbf{H}_{2\infty}$), between these two retinal planes. This homography can be readily recovered from the previous analysis as follows. Expliciting Eq. (11.19), we obtain

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \mathbf{H} & \mathbf{h} \\ \times & \times \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{1\infty} - \mathbf{e}'_1 \mathbf{r}_\infty^T & \mathbf{e}'_1 \\ \times & \times \end{bmatrix},$$

where $\mathbf{e}'_1$ is the epipole, in the view corresponding to the second position of the first camera, with respect to the view corresponding to the first position of this camera (see figure 11.10). Therefore $\mathbf{H}_{1\infty} = \mathbf{H} + \mathbf{h}\mathbf{r}_\infty^T$. If several such homographies can be estimated through a combination of motions of the rig, one can obtain the intrinsic parameters of the left camera using the techniques described in Sec. 12.2 and also the motion parameters, noted $\mathbf{R}_1$ and $\mathbf{t}_1$.

11.5.2 Euclidean calibration

This section is complementary to the previous one where we showed how to compute the monocular homographies $\mathbf{H}_{1\infty}$ (resp. $\mathbf{H}_{2\infty}$) from the epipolar geometry of the stereo rig. In this section, we show that from the displacement $\boldsymbol{\mathcal{D}}_1$ of the left camera and the displacement $\boldsymbol{\mathcal{D}}_2$ of the right camera, one can obtain the displacement $\boldsymbol{\mathcal{D}}$ between the cameras of the rig [225].

The difficulty arises from the fact that $\boldsymbol{\mathcal{D}}_1$ and $\boldsymbol{\mathcal{D}}_2$ are known in different coordinates systems. To cope with this problem, the idea is to use the commutativity of the following diagram which reflects figure 11.10:

$$\begin{array}{cccc} & \mathcal{E}_1^2 & \xrightarrow{\boldsymbol{\mathcal{D}}} & \mathcal{E}_2^2 & \\ \boldsymbol{\mathcal{D}}_1 & \uparrow & & \uparrow & \boldsymbol{\mathcal{D}}_2 \\ & \mathcal{E}_1^1 & \xrightarrow{\boldsymbol{\mathcal{D}}} & \mathcal{E}_2^1 & \end{array}$$

In this diagram, the relative displacement from the first to the second camera does not change between the initial and the final position of the stereo rig. We can thus write the matrix equation:

$$\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{D}}_1 = \boldsymbol{\mathcal{D}}_2\boldsymbol{\mathcal{D}} \tag{11.20}$$

where $\mathcal{D}$ is the 4×4 unknown matrix of the displacement from the first camera to the second camera, $\mathcal{D}_1$, $\mathcal{D}_2$ are the 4×4 displacement matrices of the first and of the second cameras, in their initial respective coordinate systems. Equation (11.20) can be decomposed in the following two matrix equations:

$$\mathbf{R}\mathbf{R}_1 = \mathbf{R}_2\mathbf{R} \tag{11.21}$$

$$(\mathbf{I} - \mathbf{R}_2)\mathbf{t} = \mu_2\mathbf{t}_2 - \mu_1\mathbf{R}\mathbf{t}_1 \tag{11.22}$$

where μ_1 and μ_2 are unknown scale factors associated to $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. The equation (11.21) has been much studied in the framework of hand-eye calibration [62] [326], [389], [58].

To solve it we use a quaternion representation of the rotations:

$$\begin{aligned} \mathbf{q}_\mathbf{R} &= (s, \mathbf{v}) \\ \mathbf{q}_{\mathbf{R}_1} &= (s_1, \mathbf{v}_1) \\ \mathbf{q}_{\mathbf{R}_2} &= (s_2, \mathbf{v}_2) \end{aligned}$$

Writing equation (11.21) with this representation yields:

$$\mathbf{q}_\mathbf{R} * \mathbf{q}_{\mathbf{R}_1} - \mathbf{q}_{\mathbf{R}_2} * \mathbf{q}_\mathbf{R} = 0 \tag{11.23}$$

where $*$ indicates the quaternion product. There exists a 4×4 matrix $\mathbf{G}$, such that:

$$\mathbf{q}_\mathbf{R} * \mathbf{q}_{\mathbf{R}_1} - \mathbf{q}_{\mathbf{R}_2} * \mathbf{q}_\mathbf{R} = \mathbf{G}\mathbf{q}_\mathbf{R}$$

A closed form solution can be obtained with two equations obtained by two displacements of the stereo rig. If we use more displacements, we can improve the results by using a linear least-squares procedure.

We now suppose that we have computed $\mathbf{R}$, as previously described. A geometrical analysis shows that the matrix $\mathbf{I} - \mathbf{R}_2$ maps all vectors in the plane perpendicular to the axis $\mathbf{r}_2$ of the rotation $\mathbf{R}_2$. Thus, starting from relation (11.22), we can write:

$$\mathbf{r}_2.(\mu_2\mathbf{t}_2 - \mu_1\mathbf{R}\mathbf{t}_1) = 0$$

This allows us to determine the ratio $a = \mu_1/\mu_2$. It is then possible to recover the direction $\mathbf{t}_\perp$ of the component of $\mathbf{t}$ orthogonal to $\mathbf{u}_2$, yielding the constraint:

$$\mathbf{t} \in \langle \mathbf{t}_\perp, \mathbf{r}_2 \rangle \tag{11.24}$$

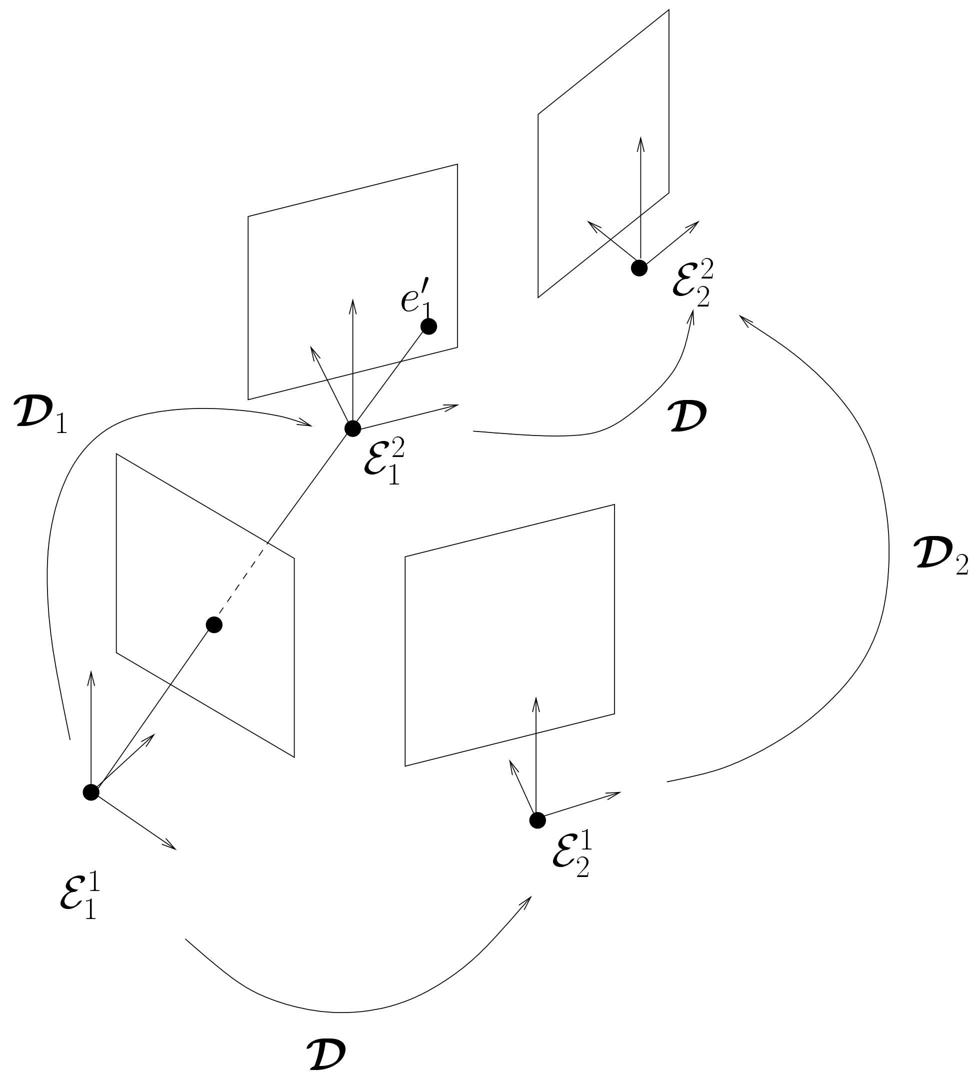


Figure 11.10: The stereo rig moves rigidly: the motion $\boldsymbol{\mathcal{D}}_1$ of the first camera, $\boldsymbol{\mathcal{D}}_2$ of the second and the relative displacement $\boldsymbol{\mathcal{D}}$ between the two cameras of the rig satisfy the constraint (11.20).

If a second movement, *for which the axis* $\mathbf{r}'_2$ *of the rotation is different*, is performed, we can compute similarly a direction $\mathbf{t'}_\perp$. Combining the two constraints (11.24), and the same with primes, we obtain $\mathbf{t}$ up to a scale factor by:

$$\mathbf{t} = \lambda(\mathbf{t}_\perp \times \mathbf{u}_2) \times (\mathbf{t'}_\perp \times \mathbf{u}'_2) \tag{11.25}$$

If we perform more than two displacements, the direction of $\mathbf{t}$ can be recovered by a linear least-squares procedure.

This completes the computation of the relative position of the two cameras, up to a scale factor.

11.6 References and further reading

Multi-view constraints and representations The quadrifocal tensor was discovered by Triggs [376] and used for reconstruction by Heyden [154, 155] who gave an algorithm for computation and extraction of projection matrices. This was improved by Hartley [144] by minimizing algebraic error.

Relations between different Trifocal tensors are described in [320], and between homographies in [428]. The relations between different multilinear constraints and the question of the minimal number of constraints necessary to generate all of them is investigated by Heyden and Astrom [158] using ideal theory. In particular, they give a precise answer to the question whether "the trilinear constraints follow from the bilinear one in the generic case".

The idea of using certain projection matrices as canonical representations for the geometries of multiple views in a stratified context is due to Luong and Vieville [227]. Heyden and Astrom [160, 157] have proposed a scheme in the same spirit, which differs by the fact that they choose four corresponding points for the canonical basis, which make relations much simpler, but dependant on the choice of these points.

Reconstruction from N views The first incremental method to build global representations (ie projection matrices) from local representations (ie Fundamental matrices and infinity homographies) in the projective and affine case are due to Luong and Vieville [227]. Avidan and Shashua [14] propose a similar approach where consecutive Fundamental matrices are connected through a property of the Trifocal tensor. So far the most general approach is that of Triggs [379] who points out to a set of linear constraints (called the closure constraints) make it possible to recover the projection matrices from the four categories of tensors, epipoles, Fundamental matrices, Trifocal tensors and quadrifocal tensors. the Trifocal tensor.

Factorization methods were introduced by Tomasi and Kanade [365] (Debrunner and Ahuja [83] presented a less general version which assumed fixed-axis rotation), originally as a batch approach for a single motion, and in the orthographic case. It has been extended to deal with weak perspective [408] in a recursive formulation, para-perspective [270], multiple motions [68], lines [284], points, lines, planes simultaneously and taking into account uncertainty [254]. In order to apply the factorization method to the projective case, the main difficulty is to recover the projective depths. Sturm and Triggs [352, 377] first discovered a way to do so, based on computation of the Fundamental matrices between pairs of views. In [163], an iterative version is given. Ueshiba and Tomita [390] propose an approach which does not rely on the computation of the Fundamental matrices. Instead, the idea is to define a measure of "rank 4 proximity" of the measurement matrix and use it to estimate the projective depths by non-linear minimization, using the affine camera model to initialize. They also derive metric constraints in case the intrinsic parameters are available.

The Bundle adjustment method has been known in the Euclidean context for a while in the photogrammetry community [47, 48, 95, 336, 3, 128, 335, 129]. It has been more recently adapted to the projective framework [251, 134, 358]. Christy and Horaud [63] provides an iterative method to solve the bundle adjustment equations in the calibrated case. The method uses the relation between perspective and weak perspective. It requires only a few affine iterations, which are linear. A comparison between the results obtained with Euclidean and projective bundle adjustment is due to Oliensis and Govindu. They find that while Euclidean is sometimes as accurate as projective even with uncertain intrinsic parameters, projective has less trouble with local minima.

Other recent work on projective reconstruction include [212, 59] (iterative methods) [264] (quasi-linear, assuming motions are small) [121, 304] (finding the point correspondences is part of the problem). Strategies to cope with degeneracies arising from non-parallax situations (fixed camera or planar structure) when recovering structure and motion in image sequences are studied in [368]. Quan et al [282] present a theoretical analysis of mimimal data necessary for reconstruction [282] when points are not seen in all views.

Cross and Zisserman [73] study the reconstruction of quadrics from their outlines. With two views, there is a one-parameter family of solutions, which is reduced to two given a point correspondence. The ambiguity is reduced for degenerate quadrics. Generally, wth three views and more, there is no ambiguity. The reconstruction of curves is tacked by Berthilsoon et al [29].

Stereo rigs The idea that with multiple views, between which there are Euclidean constraints such as those of a fixed stereo rig, can be used to recover Euclidean information, has been discovered in the context of object recognition by Forsyth et al [117].

The possibility of self-calibrating stereo rig based only on its rigidity and one general motion using Eq. 11.19 was discovered independantly by Zisserman et al [442] and Devernay and Faugeras [89]. The method in [442] requires the computation of Fundamental matrices associated with the left and right camera, in addition to the one associated with the stereo rig, and is not easily extended to deal with multiple motions. The method in [89] requires non-linear minimization to find the Euclidean parameters in $\mathcal{H}_{PE}$. The method that we presented in this chapter was based on the work of Horaud et al [166]. The work of Csurka et al [75] differs from it by proposing a closed form solution. They give parameterizations of the matrix $\mathcal{H}_{12}$ using a real Jordan factorization in the case of general motion and planar motion, which make explicit the camera intrinsic parameters. They find that four intrinsic parameters can be recovered from a single general motion (which was expected), and that three intrinsic parameters can be recovered from a single planar motion. All these methods require the estimation of a homography of $\mathbb{P}^3$ between two set of 3-D points. Csurka et al [74] present linear, non-linear, and robust methods to perform this task.

Zhang et al [438] assume that left-to-right correspondences, as well as temporal correspondences are available, so that the self-calibration problem is formulated as a Euclidean parameter minimization problem. Enciso and Vieville [93] generalize this approach. Updating camera calibration is much easier with a stereo rig [439]: if the stereo rig is initially calibrated, and at least six left-to-right and temporal correspondences are available, then one just reconstructs the points at the first instant, and then uses them with the points observed at the second instant to obtain the new projection matrices.

Brooks et al [45] consider a particular kind of stereo head, with vergence and coplanar optical axes, undergoing a motion in which the baseline remains confined to a plane. Using an Euclidean parameterization, they derive a closed form solution for self-calibration. Planar motion of a stereo rig was used by Beardsley et al [23]. who show that affine reconstruction can be obtained. Horaud, Czurka, Demirdjian, and Ruf have done a lot of work on uncalibrated stereo rigs, using a Jordan factorization of the homography $\mathcal{H}$ to find an appropriate parameterization. They study in great detail the pure rotation of a stereo rig [302] on a pan-tilt head, and the pure translation of a stereo rig [301]. In the latter case, a closed form solution is given for the decomposition of $\mathcal{H}$ into the three parameters of the plane at infinity, and the translation parameters. The analysis is extended to rigid and articulated motions as seen by a stereo rig in [303]. In [87], they propose to segment rigid motions based on the homography $\mathcal{H}$ using a stereo rig, extending the idea of segmentation of rigid motions based on the Fundamental matrix using a single camera.