Computer Vision
A Modern Approach

David A. Forsyth
University of California at Berkeley

Jean Ponce
University of Illinois at Urbana-Champaign

An Alan R. Apt Book

Prentice Hall
Upper Saddle River, New Jersey 07458

13

Projective Structure from Motion

This chapter addresses once again the recovery of scene structure and/or camera motion from correspondences established by matching the images of n points in m pictures. This time, however, we assume a perspective projection model. Given n *fixed* points P_j $(j = 1, \dots, n)$ observed by m cameras and the corresponding mn *homogeneous* coordinate vectors $\boldsymbol{p}_{ij} = (u_{ij}, v_{ij}, 1)^T$ of their images, let us write the corresponding perspective projection equations as

$$\begin{cases} u_{ij} = \dfrac{\boldsymbol{m}_{i1} \cdot \boldsymbol{P}_j}{\boldsymbol{m}_{i3} \cdot \boldsymbol{P}_j} \\ v_{ij} = \dfrac{\boldsymbol{m}_{i2} \cdot \boldsymbol{P}_j}{\boldsymbol{m}_{i3} \cdot \boldsymbol{P}_j} \end{cases} \quad \text{for} \quad i = 1, \dots, m \quad \text{and} \quad j = 1, \dots, n, \tag{13.1}$$

where $\boldsymbol{m}_{i1}^T$, $\boldsymbol{m}_{i2}^T$, and $\boldsymbol{m}_{i3}^T$ denote the rows of the 3×4 projection matrix $\mathcal{M}_i$ associated with camera number i in some fixed coordinate system, and $\boldsymbol{P}_j$ denotes the *homogeneous* coordinate vector of the point P_j in that coordinate system. We define *projective structure from motion* as the problem of estimating the m matrices $\mathcal{M}_i$ and the n vectors $\boldsymbol{P}_j$ from the mn image correspondences $\boldsymbol{p}_{ij}$.

When $\mathcal{M}_i$ and $\boldsymbol{P}_j$ are solutions of Eq. (13.1), so are of course $\lambda_i \mathcal{M}_i$ and $\mu_j \boldsymbol{P}_j$ for any nonzero values of λ_i and μ_j. In particular, as already noted in chapter 2, the matrices $\mathcal{M}_i$ satisfying Eq. (13.1) are only defined up to scale, with 11 independent parameters, and so are the vectors $\boldsymbol{P}_j$, with 3 independent parameters (when necessary, these can be reduced to the canonical form $(x_j, y_j, z_j, 1)^T$ as long as their fourth coordinate is not zero, which is the generic case). Like its affine cousin, projective structure from motion suffers from a deeper ambiguity that justifies its name: When the camera calibration parameters are unknown, the projection matrices $\mathcal{M}_i$ are, according to Theorem 1 (chapter 2), arbitrary rank-3 3×4 matrices. Hence, if $\mathcal{M}_i$ and

$\boldsymbol{P}_j$ are solutions of Eq. (13.1), so are $\mathcal{M}_i' = \mathcal{M}_i\mathcal{Q}$ and $\boldsymbol{P}_j' = \mathcal{Q}^{-1}\boldsymbol{P}_j$, where $\mathcal{Q}$ is a *projective transformation matrix* (i.e., an arbitrary nonsingular 4×4 matrix). The matrix $\mathcal{Q}$ is only defined up to scale, with 15 free parameters, since multiplying it by a nonzero scalar simply amounts to applying inverse scalings to $\mathcal{M}_i$ and $\boldsymbol{P}_j$. Since Eq. (13.1) provides $2mn$ constraints on the $11m$ parameters of the matrices $\mathcal{M}_i$ and the $3n$ parameters of the vectors $\boldsymbol{P}_j$, taking into account the *projective ambiguity* of structure from motion suggests that this problem admits a finite number of solutions as soon as $2mn \geq 11m + 3n - 15$. For $m = 2$, seven point correspondences should thus be sufficient to determine (up to a projective transformation) the two projection matrices and the position of any other point. This is confirmed formally in Sections 13.2 and 13.3.

In the rest of this chapter, *projective geometry* plays the role that affine geometry played in chapter 12, and it affords a similar overall methodology. Once again, ignoring (at first) the Euclidean constraints associated with calibrated cameras allows us to linearize the recovery of the *projective* scene structure and camera motion from point correspondences. We then exploit the geometric constraints associated with (partially or fully) calibrated perspective cameras to upgrade the projective reconstruction to a Euclidean one.

13.1 ELEMENTS OF PROJECTIVE GEOMETRY

The means of measurement available in projective geometry are even more primitive than those available in affine geometry. The affine notion of ratios of lengths along parallel lines and, in fact, the notion of parallelism are gone. The concepts of points, lines, and planes remain, however, as well as a new, weaker scalar measure of the arrangement of collinear points—the *cross-ratio*. As in the affine case, a rigorous axiomatic introduction to projective geometry would be out of place in this book, and we remain rather informal in the rest of this section.

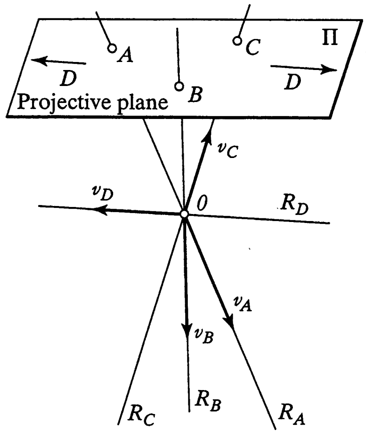
13.1.1 Projective Spaces

Let us consider a real vector space $\vec{X}$ of dimension $n + 1$. If $\boldsymbol{v}$ is a nonzero element of $\vec{X}$, the set $\mathbb{R}\boldsymbol{v}$ of all vectors proportional to $\boldsymbol{v}$ is called a *ray*, and it is uniquely characterized by any one of its nonzero elements. The *real projective space* $X = P(\vec{X})$ of dimension n associated with $\vec{X}$ is the set of rays in $\vec{X}$ or, equivalently, the quotient of the set $\vec{X}\backslash 0$ of nonzero vectors in $\vec{X}$ under the equivalence relation "$\boldsymbol{v} \sim \boldsymbol{v}'$ if and only if $\boldsymbol{v} = k\boldsymbol{v}'$ for some $k \in \mathbb{R}$". Elements of X are called *points*, and we say that a family of points are linearly dependent (resp. independent) when representative vectors for the corresponding rays are linearly dependent (resp. independent). The map $p : \vec{X}\backslash 0 \to P(\vec{X})$ associates with any nonzero element $\boldsymbol{v}$ of $\vec{X}$ the corresponding point $p(\boldsymbol{v})$ of X.

Example 13.1 A Model of $P(\mathbb{R}^3)$.

Consider an affine plane Π of $\mathbb{R}^3$. The rays of $\mathbb{R}^3$ that are not parallel to Π are in one-to-one correspondence with the points of this plane. For example, the rays R_A, R_B, and R_C associated with the vectors $\boldsymbol{v}_A$, $\boldsymbol{v}_B$, and $\boldsymbol{v}_C$ below can be mapped onto the points A, B and C where they intersect Π. The vectors $\boldsymbol{v}_A$, $\boldsymbol{v}_B$, and $\boldsymbol{v}_C$ are linearly independent, and so are (by definition) the points A, B, and C.

As a ray gets close to being parallel to Π, the point where it intersects this plane recedes to infinity, and in fact it can be shown that a model of the projective plane $P(\mathbb{R}^3)$ (i.e., a projective space $\hat{\Pi}$ of dimension 2 isomorphic to $P(\mathbb{R}^3)$) can be constructed by adding to Π a one-dimensional set of *points at infinity* associated with the rays parallel to this plane. Here, for example, the ray R_D parallel to Π maps onto the point at infinity D of $\hat{\Pi}$.



Since any affine plane can be mapped onto $\mathbb{R}^2$ by choosing some affine coordinate system, Example 13.1 suggests that affine planes, and for that matter $\mathbb{E}^3$ or any other affine space, can somehow be embedded in projective spaces, an appropriate choice of points at infinity completing the embedding. Such a completion process is presented in Section 13.1.3.

13.1.2 Projective Subspaces and Projective Coordinates

Consider an $(m+1)$-dimensional vector subspace $\vec{Y}$ of $\vec{X}$. The set Let $Y = P(\vec{Y})$ of rays in $\vec{Y}$ is called a *projective subspace* of X, and its *dimension* is m. Given a basis $(\boldsymbol{e}_0, \boldsymbol{e}_1, \dots, \boldsymbol{e}_m)$ for $\vec{Y}$, we can associate with each point P in Y a one-parameter family of elements of $\mathbb{R}^{m+1}$—namely, the coordinate vectors $(x_0, x_1, \dots, x_m)^T$ of the vectors $\boldsymbol{v} \in \vec{Y}$ such that $P = p(\boldsymbol{v})$. These tuples are proportional to one another, and a representative tuple is called a set of *homogeneous projective coordinates* of the point P.

Homogeneous coordinates can also be characterized intrinsically in terms of families of points in Y: Consider $m+1$ $(m \le n)$ linearly independent points $A_0, A_1, \dots, A_m$ and $m+1$ vectors $\boldsymbol{v}_i$ $(i = 0, 1, \dots, m)$ representative of the corresponding rays. If an additional point A^* linearly depends on the points A_i and $\boldsymbol{v}^*$ is a representative vector of the corresponding ray, we can write

$$\boldsymbol{v}^* = \mu_0 \boldsymbol{v}_0 + \mu_1 \boldsymbol{v}_1 + \dots + \mu_m \boldsymbol{v}_m.$$

The coefficients μ_i are not uniquely determined since *each* vector $\boldsymbol{v}_i$ is only defined up to a nonzero scale factor. However, when none of the coefficients μ_i vanishes (i.e., when $\boldsymbol{v}^*$ does not lie in the vector subspace spanned by any m vectors $\boldsymbol{v}_i$ or, equivalently, when the corresponding points are linearly independent), we can uniquely define the $m+1$ nonzero vectors $\boldsymbol{e}_i = \mu_i \boldsymbol{v}_i$ such that

$$\boldsymbol{v}^* = \boldsymbol{e}_0 + \boldsymbol{e}_1 + \dots + \boldsymbol{e}_m.$$

In particular, any vector $\boldsymbol{v}$ linearly dependent on the vectors $\boldsymbol{v}_i$ can now be written *uniquely* as

$$\boldsymbol{v} = x_0 \boldsymbol{e}_0 + x_1 \boldsymbol{e}_1 + \dots + x_m \boldsymbol{e}_m.$$

This defines a one-to-one correspondence between the rays $\mathbb{R}(x_0, x_1, \dots, x_m)^T$ of $\mathbb{R}^{m+1}$ and a projective subspace S_m of X. S_m is, in fact, the projective space Y associated with the vector subspace $\vec{Y}$ of $\vec{X}$ spanned by the vectors $\boldsymbol{v}_i$ (or, equivalently, by the vectors $\boldsymbol{e}_i$). If $P = p(\boldsymbol{v})$ is the point of S_m associated with the ray $\mathbb{R}\boldsymbol{v}$, the numbers $x_0, x_1, \dots, x_m$ are called the homogeneous (projective) coordinates of P in the *projective coordinate system* determined by the $m+1$ *fundamental points* A_i and the *unit point* A^*. Note that, since the vector $\boldsymbol{v}$ associated with a ray is only defined up to scale, so are the homogeneous coordinates of a point.

It is a simple matter to verify that the coordinate vectors of the fundamental and unit points in the corresponding projective frame have a particularly simple form—namely,

$$A_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad A_m = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \quad \text{and } A^* = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

It should be clear that the two notions of homogeneous coordinates that have been introduced in this section coincide. The only difference is in the choice of the coordinate vectors $\boldsymbol{e}_0, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$, that are given a priori in the former case and constructed from the points forming a given projective frame in the latter one.

Example 13.2 Projective Coordinate Changes.

Given some coordinate system $(A) = (A_0, A_1, A_2, A_3, A^*)$ for the three-dimensional projective space X, we can define the (homogeneous projective) coordinate vector of any point P as ${}^A P = ({}^A x_0, {}^A x_1, {}^A x_2, {}^A x_3)^T$. Let us now consider a second projective frame $(B) = (B_0, B_1, B_2, B_3, B^*)$ for X. It can easily be shown (see Exercises) that the corresponding change of coordinates can be written as

$$\rho\, {}^B P = {}^B_A\mathcal{T}\, {}^A P, \tag{13.2}$$

where ${}^B_A\mathcal{T}$ is a 4×4 projective transformation matrix defined up to scale, and ρ is a scalar chosen so the scales of the two sides of the equations are the same. Let us now show how to compute this matrix. Writing Eq. (13.2) for the points defining the frame (A) yields

$$\rho_0\, {}^B A_0 = {}^B_A\mathcal{T} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \rho_1\, {}^B A_1 = {}^B_A\mathcal{T} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \rho_2\, {}^B A_2 = {}^B_A\mathcal{T} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \rho_3\, {}^B A_3 = {}^B_A\mathcal{T} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

and

$$\rho^*\, {}^B A^* = {}^B_A\mathcal{T} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Since the matrix ${}^B_A\mathcal{T}$ is only defined up to a scale factor, we can choose $\rho^* = 1$, and it follows that

$${}^B_A\mathcal{T} = \begin{pmatrix} \rho_0\, {}^B A_0 & \rho_1\, {}^B A_1 & \rho_2\, {}^B A_2 & \rho_3\, {}^B A_3 \end{pmatrix},$$

where the scalars ρ_i are the solutions of the linear system

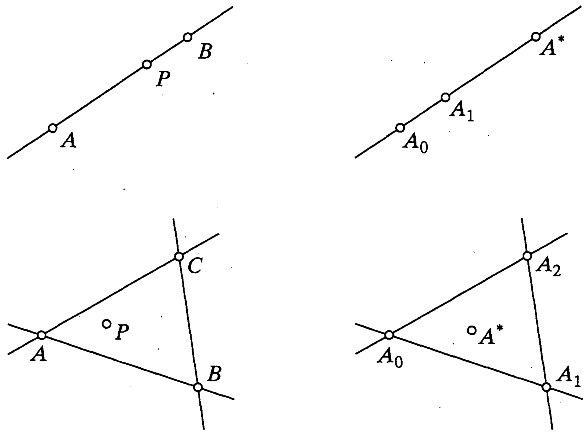
$$\begin{pmatrix} {}^B A_0 & {}^B A_1 & {}^B A_2 & {}^B A_3 \end{pmatrix} \begin{pmatrix} \rho_0 \\ \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} = {}^B A^*.$$

Note the obvious similarity with the formulas for changes of Euclidean or affine coordinate systems in chapters 2 and 12. Similar formulas for coordinate changes can be written for arbitrary projective spaces of finite dimension.

A projective subspace S_1 of dimension 1 of X is called a *line*. Linear subspaces of dimension 2 and $n-1$ are, respectively, called *planes* and *hyperplanes*. A hyperplane S_{n-1} consists of the set of points P linearly dependent on n linearly independent points $P_0, P_1, \dots, P_{n-1}$.

Example 13.3 Projective lines and planes.

A projective line is uniquely determined by two distinct points A and B lying on it, but defining a projective frame requires three distinct points A_0, A_1, and A^* on that line. Likewise, a plane is uniquely determined by three points A, B, and C lying in it, but defining a projective frame requires four points in that plane: Three fundamental points A_0, A_1, and A_2 forming a nondegenerate triangle and a unit point A^* not lying on one of the edges of this triangle.



However, if $\boldsymbol{A}$ and $\boldsymbol{B}$ denote the coordinate vectors of two distinct points in some projective frame for the line passing through these points, the coordinate vector $\boldsymbol{P}$ of any point on that line can be written uniquely as $\boldsymbol{P} = \lambda\boldsymbol{A} + \mu\boldsymbol{B}$. This follows immediately from the fact that the rays R_A and R_B associated with distinct points A and B are linearly independent, but the ray R_P associated with a point P on the same line lies in the vector plane defined by R_A and R_B. Likewise, if $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ denote the coordinate vectors of three noncollinear points in some projective frame for the plane they lie in, the coordinate vector $\boldsymbol{P}$ of any point in that plane can be written uniquely as $\boldsymbol{P} = \lambda\boldsymbol{A}+\mu\boldsymbol{B}+\nu\boldsymbol{C}$.

13.1.3 Affine and Projective Spaces

Example 13.1 introduced (informally) the idea of embedding an affine plane into a projective one with the addition of a one-dimensional set of points at infinity. More generally, it is possible to construct the *projective closure* $\tilde{X}$ of an affine space X of dimension n by adding to it a set of points at infinity associated with the directions of its lines. These points form a hyperplane of $\tilde{X}$ called the *hyperplane at infinity* and denoted by ∞_X.

Let us pick some point A in X and introduce $\tilde{X} \stackrel{\text{def}}{=} P(\vec{X} \times \mathbb{R})$, where $\vec{X}$ is the vector space underlying X. We can embed X into $\tilde{X}$ via the injective map $J_A : X \to \tilde{X}$ defined by $J_A(P) = p(\overrightarrow{AP}, 1)$ (Figure 13.1).[1] The complement of $J_A(X)$ in $\tilde{X}$ is the hyperplane at infinity $\infty_X \stackrel{\text{def}}{=} P(\vec{X} \times \{0\})$ mentioned earlier.

[1]Here we identify X and the underlying vector space $\vec{X}$ by identifying each point P in X with the vector $\overrightarrow{AP}$. This *vectorialization* process is of course dependent on the choice of the origin A, but it can easily be shown that $\tilde{X}$ is indeed independent of that choice. A more rigorous approach to the projective completion process is to introduce the *universal vector space* associated with an affine space, but it would be out of place here. See Berger (1987, chapter 5) for details. Note also the abuse of notation in writing $p(\nu, \lambda)$ for $p((\nu, \lambda))$.

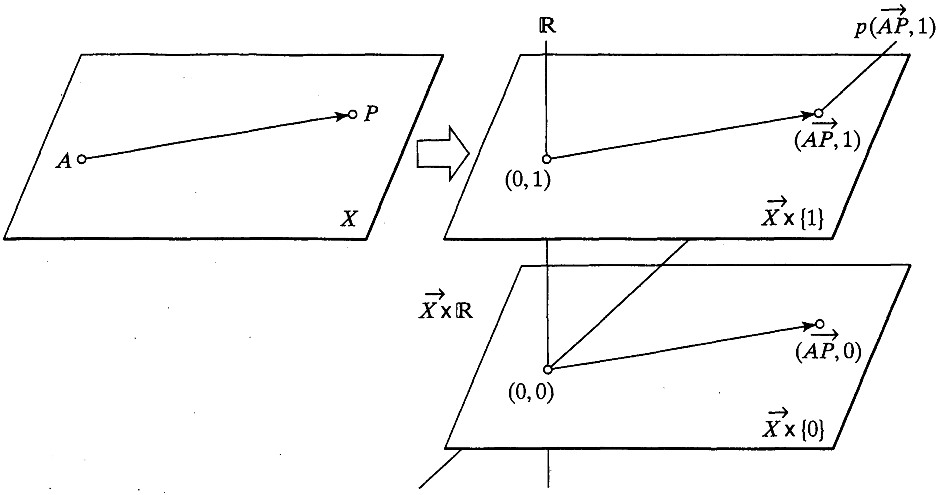


Figure 13.1 The projective completion of an affine space.

Now consider a fixed affine frame $(A_0, A_1, \ldots, A_n)$ of X and embed X into $\tilde{X}$ using J_{A_0}. The vectors $\overrightarrow{A_0A_i}$ $(i = 1, \ldots, n)$ form a basis of $\vec{X}$, thus the $n+1$ vectors $\boldsymbol{e}_i \stackrel{\text{def}}{=} (\overrightarrow{A_0A_i}, 0)$ $(i = 1, \ldots, n)$ and $\boldsymbol{e}_{n+1} \stackrel{\text{def}}{=} (\mathbf{0}, 1)$ form a basis of $\vec{X} \times \mathbb{R}$. In particular, if $(x_1, \ldots, x_n)$ denote the affine coordinates of P in the basis $(A_0, A_1, \ldots, A_n)$ of X, we have

$$\begin{aligned} J_{A_0}(P) &= p(\overrightarrow{A_0P}, 1) = p(x_1\overrightarrow{A_0A_1} + \cdots + x_n\overrightarrow{A_0A_n}, 1) \\ &= p(x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n + \boldsymbol{e}_{n+1}), \end{aligned}$$

and the homogeneous projective coordinates of $J_{A_0}(P)$ associated with the basis of $\vec{X} \times \mathbb{R}$ formed by the vectors $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n+1})$ are thus $(x_1, \ldots, x_n, 1)$. The coordinates of points in ∞_X, on the other hand, have the form $(x_1, \ldots, x_n, 0)$. In particular, the projective completion process justifies, at long last, the representation of image and scene points by homogeneous coordinates introduced in chapter 2 and used throughout this book.

The introduction of points at infinity frees projective geometry from the numerous exceptions encountered in the affine case. For example, parallel lines in some affine plane Π do not intersect unless they coincide. In contrast, any two distinct lines in a projective plane intersect in exactly one point (this is because the associated vector spaces intersect along a ray), with pairs of parallel lines in Π intersecting at the point at infinity in $\tilde{\Pi}$ that is associated with their common direction (see Exercises).

13.1.4 Hyperplanes and Duality

As mentioned before, two distinct lines of a projective plane have exactly one common point. Likewise, two distinct points belong to exactly one line. These two statements can actually be taken as *incidence axioms*, leading to a purely axiomatic construction of the projective plane. Points and lines play a symmetric or, more precisely, *dual* role in these statements.

To introduce *duality* a bit more generally, let us equip the n-dimensional projective space X with a fixed projective frame and consider $n+1$ points $P_0, P_1, \ldots, P_n$ lying in some hyperplane S_{n-1} of X. Since these points are by construction linearly dependent, the $(n+1) \times (n+1)$ matrix formed by collecting their coordinate vectors is singular. Expanding the determinant of this matrix with respect to its last column yields

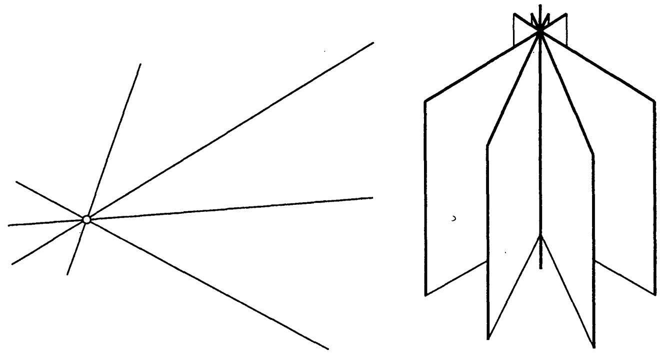
$$u_0x_0 + u_1x_1 + \cdots + u_nx_n = 0, \tag{13.3}$$

where $(x_0, x_1, \dots, x_n)$ denote the *homogeneous* coordinates of P_n and $(u_0, u_1, \dots, u_n)$ are functions of the coordinates of the points $P_0, P_1, \dots, P_{n-1}$. Note that we have refrained to set the last coordinate of the point P_n to 1 here to emphasize the symmetry between the scalars u_i and x_i.

Equation (13.3) is satisfied by every point P_n in the hyperplane S_{n-1}, and it is called the equation of S_{n-1} (note the similarity with the affine case). Conversely, it is easily shown that any equation of the form in Eq. (13.3) where at least one of the coefficients u_i is nonzero is the equation of some hyperplane. Since the coefficients u_i in Eq. (13.3) are only defined up to some common scale factor, there exists a one-to-one correspondence between the rays of $\mathbb{R}^{n+1}$ and the hyperplanes of X, and it follows that we can define a second projective space $X^* = P(\vec{X}^*)$ formed by these hyperplanes and called the *dual* of $\vec{X}$ (this is justified by the fact that X^* can be shown to be the projective space associated with the dual vector space $\vec{X}^*$ of $\vec{X}$). The scalars $(u_0, u_1, \dots, u_n)$ define homogeneous projective coordinates for the point corresponding to the hyperplane S_{n-1} in X^*, and Eq. (13.3) can also be seen as defining the set of hyperplanes passing through the point P_n.

Example 13.4 The Dual of a Line.

Points and lines are dual notions in $\mathbb{P}^2 \stackrel{\text{def}}{=} \tilde{\mathbb{E}}^2$, points and planes are dual in $\mathbb{P}^3 \stackrel{\text{def}}{=} \tilde{\mathbb{E}}^3$, but points and lines are *not* dual in $\mathbb{P}^3$. In general, what is the dual of a line in X? A line is a one-dimensional linear subspace of X whose elements are linearly dependent on two points on the line. Likewise, a line of X^* is a one-dimensional subspace of the dual, called a *pencil of hyperplanes*, whose elements are linearly dependent on two hyperplanes in the pencil. In the plane, the dual of a line is a pencil of lines intersecting at a common point.



In three dimensions, the dual of a line is a pencil of planes that intersect along a common line.

Let us close this section by noting (without proof) that any geometric theorem that holds for points in X induces a corresponding theorem for hyperplanes (i.e., points in X^*) and vice versa, the two theorems being said to be dual of each other.

13.1.5 Cross-Ratios and Projective Coordinates

This section focuses on the three-dimensional projective space $\tilde{\mathbb{E}}^3$. The *nonhomogeneous* projective coordinates of a point can be defined geometrically in terms of *cross-ratios*. In the affine case, given four collinear points A, B, C, D such that A, B, and C are distinct, we define the cross-ratio of these points as

$$\{A, B; C, D\} \stackrel{\text{def}}{=} \frac{\overline{CA}}{\overline{CB}} \times \frac{\overline{DB}}{\overline{DA}},$$

where $\overline{PQ}$ denotes the signed distance between two points P and Q for some choice of orientation of the line Δ joining them. The orientation of this line is fixed but arbitrary since reversing it obviously does not change the cross-ratio. Note that $\{A, B; C, D\}$ is, a priori, only defined when $D \neq A$ since its calculation involves a division by zero when $D = A$. We extend the definition of the cross-ratio to the whole affine line by using the symbol ∞ to denote the ratio formed by dividing any nonzero real number by zero and to the whole projective line $\tilde{\Delta}$ by defining $\{A, B; C, \infty_\Delta\} = \overline{CA}/\overline{CB}$. Alternatively, given three points A, B, and C on a *projective* line Δ, it can be shown that there exists a unique projective transformation $h : \Delta \to \tilde{\mathbb{R}}$ mapping Δ onto the projective completion $\tilde{\mathbb{R}} = \mathbb{R} \cup \infty$ of the real line such that $h(A) = \infty$, $h(B) = 0$ and $h(C) = 1$. The cross-ratio can also be defined by $\{A, B; C, D\} \stackrel{\text{def}}{=} h(D)$.

Given a projective frame (A_0, A_1, A^*) for a line Δ and a point P lying on Δ with homogeneous coordinates (x_0, x_1) in that frame, we can define a nonhomogeneous coordinate for P as $k_0 = x_0/x_1$. The scalar k_0 is sometimes called *projective parameter* of P, and it is easy to show that $k_0 = \{A_0, A_1; A_2, P\}$.

As noted earlier, a set of lines passing through the same point O is called a *pencil of lines*. The cross-ratio of four *coplanar* lines Δ_1, Δ_2, Δ_3 and Δ_4 in some pencil is defined as the cross-ratio of the intersections of these lines with any other line Δ in the same plane that does not pass through O, and it is easily shown to be independent of the choice of Δ (Figure 13.2a).

Consider now four planes Π_1, Π_2, Π_3, and Π_4 in the same pencil, and denote by Δ their common line. The cross-ratio of these planes is defined as the cross-ratio of the pencil of lines formed by their intersection with any other plane Π not containing Δ (Figure 13.2b). Once again, the cross-ratio is easily shown to be independent of the choice of Π.

In the plane, the nonhomogeneous projective coordinates (k_0, k_1) of the point P in the basis (A_0, A_1, A_2, A^*) are defined by $k_0 = x_0/x_2$ and $k_1 = x_1/x_2$, and it can be shown that

$$\begin{cases} k_0 = \{A_1A_0,\ A_1A_2;\ A_1A^*,\ A_1P\}, \\ k_1 = \{A_0A_1,\ A_0A_2;\ A_0A^*,\ A_0P\}, \end{cases}$$

where MN denotes the line joining the points M and N, and $\{\Delta_1, \Delta_2; \Delta_3, \Delta_4\}$ denotes the cross-ratio of the pencil of lines Δ_1, Δ_2, Δ_3, Δ_4.

Similarly, the nonhomogeneous projective coordinates (k_0, k_1, k_2) of the point P in the basis $(A_0, A_1, A_2, A_3, A^*)$ are defined by $k_0 = x_0/x_3$, $k_1 = x_1/x_3$, and $k_2 = x_2/x_3$, and it can be

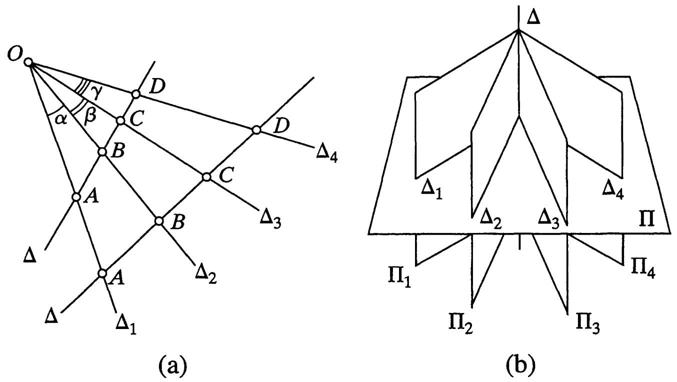


Figure 13.2 Definition of the cross-ratio of: (a) four lines, and (b) four planes. As shown in the exercises, the cross-ratio $\{A, B; C, D\}$ only depends on the three angles α, β, and γ. In particular, we have $\{A, B; C, D\} = \{A', B'; C', D'\}$.

shown that

$$\begin{cases} k_0 = \{A_1A_2A_0,\ A_1A_2A_3;\ A_1A_2A^*,\ A_1A_2P\}, \\ k_1 = \{A_2A_0A_1,\ A_2A_0A_3;\ A_2A_0A^*,\ A_2A_0P\}, \\ k_2 = \{A_0A_1A_2,\ A_0A_1A_3;\ A_0A_1A^*,\ A_0A_1P\}, \end{cases}$$

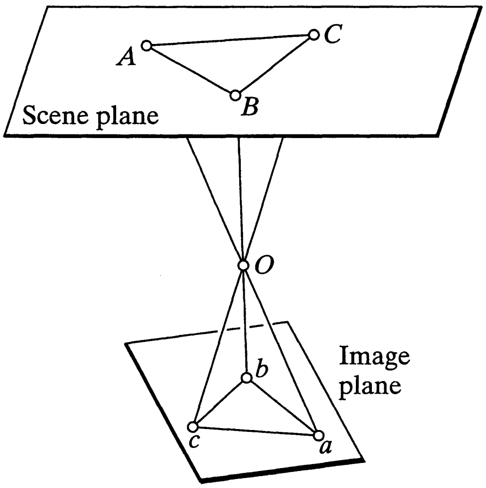
where LMN denotes the plane spanned by the three points L, M, and N, and $\{\Pi_1, \Pi_2; \Pi_3, \Pi_4\}$ denotes the cross-ratio of the pencil of planes $\Pi_1, \Pi_2, \Pi_3, \Pi_4$.

13.1.6 Projective Transformations

Consider a bijective linear map $\vec{\psi} : \vec{X} \to \vec{X}'$ between two vector spaces $\vec{X}$ and $\vec{X}'$. By linearity, $\vec{\psi}$ maps rays of $\vec{X}$ onto rays of $\vec{X}'$. Since it is bijective, it also maps nonzero vectors onto nonzero vectors, and we can define the induced map $\psi : P(\vec{X}) \to P(\vec{X}')$ by $\psi(p(v)) \stackrel{\text{def}}{=} p(\vec{\psi}(v))$ for any $v \neq 0$ in $\vec{X}$. The map ψ is bijective and is called a *projective transformation* (or *homography*). It is easy to show that projective transformations form a group under the law of composition of maps. When $\vec{X}' = \vec{X}$, this group is called the *projective group* of $X = P(\vec{X})$.

Example 13.5 Projective correspondence between coplanar points and their pictures.

Consider two planes and a point O lying outside these planes in $\mathbb{E}^3$. As shown in the exercises, the perspective projection mapping any point A in the (projective closure of the) first plane onto the intersection of the line AO with the (projective closure of the) second plane is a projective transformation.



This property should not come as a surprise since, following Example 13.1, the two (projective) planes can be thought of as models of the projective spaces associated with the set of rays through the point O.

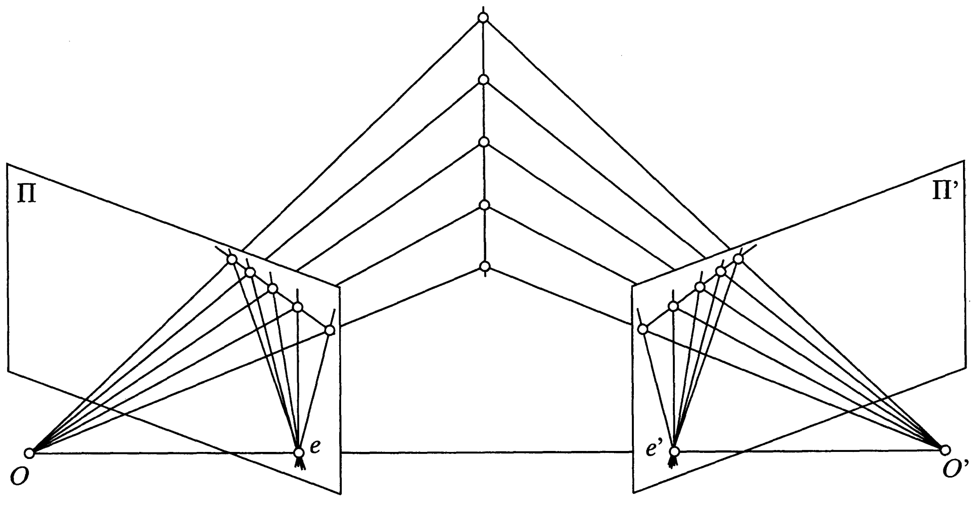
Projective geometry can be thought of as the study of the properties of projective spaces that are preserved by homographies. An example of such an *invariant* is the linear independence (or dependence) of a family of points. Given a projective transformation $\psi : X \to X'$, let us consider $m+1$ linearly independent vectors $v_0, v_1, \ldots, v_m$ in $\vec{X}$ and the corresponding points $A_0, A_1, \ldots, A_m$ in X. Since $\vec{\psi}$ is bijective, the vectors $\vec{\psi}(v_i)$ are linearly independent and so are the points $A'_i = \psi(A_i)$. It follows immediately that if $(A) = (A_0, A_1, \ldots, A_{n+1})$ is a projective frame for the n-dimensional projective space X, so is $(A') = (A'_0, A'_1, \dots, A'_{n+1})$ for X'. Conversely, given two n-dimensional projective spaces X and X', equipped respectively with the bases $(A_0, A_1, \dots, A_{n+1})$ and $(A'_0, A'_1, \dots, A'_{n+1})$, it can be shown that there exists a unique homography $\psi : X \to X'$ such that $\psi(A_i) = A'_i$ for $i = 0, 1, \dots, n+1$.

Projective coordinates form a second invariant. Indeed, due to the linearity of the underlying map $\vec{\psi}$, if the point P has coordinates $(x_0, x_1, \dots, x_n)$ in the projective frame $(A_0, A_1, \dots, A_{n+1})$ of X, the point $\psi(P)$ has the same coordinates in the coordinate frame of X' formed by the points $A'_i = \psi(A_i)$. In fact, projective transformations can be characterized as mappings that transform lines into lines and preserve cross-ratios (thus projective coordinates). Coming back to Example 13.5, it follows that an image of a set of coplanar points completely determines the projective coordinates of these points relative to the frame formed by four of them. This proves useful in designing invariant-based recognition systems in later chapters.

Like a rigid or an affine transformation, a homography ψ between two projective spaces of dimension n can conveniently be represented by an $(n+1) \times (n+1)$ matrix once coordinate systems (F) and (F') for these spaces have been chosen: This is again due to the linearity of the underlying operator $\vec{\psi}$. Thus, if $P' = \psi(P)$, we can write ${}^{F'}P' = \mathcal{Q}\,{}^{F}P$, where $\mathcal{Q}$ is a nonsingular $(n+1)\times(n+1)$ matrix defined up to scale since homogeneous projective coordinates are only defined up to scale.

Example 13.6 Parameterizing the Fundamental Matrix.

Let us revisit the problem of determining the epipolar geometry of uncalibrated cameras. This problem was introduced in chapter 10, where we gave without proof an explicit parameterization of the fundamental matrix. We now construct this parameterization. Let us define the *epipolar transformation* as the mapping from one set of epipolar lines onto the other one. As shown by the diagram, this transformation is a homography.



Indeed, the epipolar planes associated with the two cameras form a pencil whose spine is the baseline joining the two optical centers. This pencil intersects the corresponding image planes along the two families of epipolar lines, and the cross-ratio of any quadruple of lines in either family is of course the same as the cross-ratio of the corresponding planes. In turn, this means that the epipolar transformation preserves the cross-ratio and is therefore a projective transformation.

Let us denote by $(\alpha, \beta)^T$ and $(\alpha', \beta')^T$ the (affine) coordinates of the two epipoles e and e' in the corresponding image coordinate systems, and let us use $(u, v)^T$ and $(u', v')^T$ to denote the coordinates of points on matching epipolar lines l and l'. Using the fact that the linear map associated with the epipolar transformation maps the ray $\mathbb{R}(u-\alpha, v-\beta)^T$ onto the ray $\mathbb{R}(u'-\alpha', v'-\beta')^T$, it is easy to show (see Exercises) that the slopes τ and τ' of the lines l and l' satisfy

$$\tau' = \frac{a\tau + b}{c\tau + d}, \quad \text{with} \quad \tau \stackrel{\text{def}}{=} \frac{v-\beta}{u-\alpha} \quad \text{and} \quad \tau' \stackrel{\text{def}}{=} \frac{v'-\beta'}{u'-\alpha'}. \tag{13.4}$$

Clearing the denominators in Eq. (13.4) yields a bilinear expression in u, v and u', v', easily rewritten as $\boldsymbol{p}^T\mathcal{F}\boldsymbol{p}' = 0$, where $\mathcal{F}$ is written in the form given without proof in chapter 10—that is,

$$\mathcal{F} = \begin{pmatrix} b & a & -a\beta - b\alpha \\ -d & -c & c\beta + d\alpha \\ d\beta' - b\alpha' & c\beta' - a\alpha' & -c\beta\beta' - d\beta'\alpha + a\beta\alpha' + b\alpha\alpha' \end{pmatrix}.$$

13.1.7 Projective Shape

Following the approach used in the affine case, we say that two point sets S and S' in some projective space X are *projectively equivalent* when there exists a projective transformation $\psi : X \to X$ such that S' is the image of S under ψ. As in the affine case, it is easy to show that projective equivalence is an equivalence relation, and the *projective shape* of a point set S in X is defined as the equivalence class of all projectively equivalent point sets. Likewise, projective structure from motion can now be redefined as the problem of recovering the projective shape of the observed scene (and/or the equivalence classes formed by the corresponding projection matrices) from features matched in an image sequence.

13.2 PROJECTIVE STRUCTURE AND MOTION FROM BINOCULAR CORRESPONDENCES

The rest of this chapter is concerned with the recovery of the three-dimensional projective structure of a scene assuming that n points have been tracked in m images of this scene. This section focuses on the case of two images. Structure and motion estimation from three or more views are addressed in the next two sections. We assume that the epipoles are known, which, as shown in chapter 10, requires establishing at least seven point correspondences.

13.2.1 Geometric Scene Reconstruction

Let us start with a geometric method for estimating the projective shape of a scene when the epipoles are known. The inherent ambiguity of projective structure from motion simplifies our task by allowing us to choose appropriate points as a projective frame.

Let us assume that we observe four noncoplanar points A, B, C, D with a weakly-calibrated stereo rig (Figure 13.3). Let O' (resp. O'') denote the position of the optical center of the first (resp. second) camera. For any point P let p' (resp. p'') denote the position of the projection of P into the first (resp. second) image and let P' (resp. P'') denote the intersection of the ray $O'P$ (resp. $O''P$) with the plane ABC. The epipoles are e' and e'', and the baseline intersects the plane ABC in E. (Clearly, $E' = E'' = E$, $A' = A'' = A$, etc.)

We choose A, B, C, O', O'' as a basis for projective three-space, and our goal is to reconstruct the position of D. Choosing a', b', c', e' as a basis for the first image plane, we can measure the coordinates of d' in this basis and reconstruct the point D' in the basis A, B, C, E of the plane ABC. Similarly, we can reconstruct the point D'' from the projective coordinates of d'' in the basis a'', b'', c'', e'' of the second image plane. The point D is finally reconstructed as the intersection of the two lines $O'D'$ and $O''D''$.

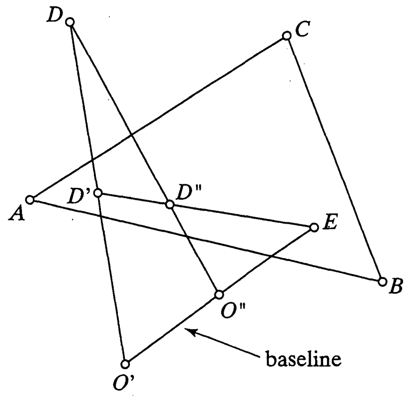


Figure 13.3 Geometric construction of the projective coordinates of the point D in the basis formed by the five points A, B, C, O', and O''.

We can now express this geometric construction in algebraic terms. It turns out to be simpler to reorder the points of our projective frame and calculate the nonhomogeneous projective coordinates of D in the basis formed by the tetrahedron A, O'', O', B and the unit point C. These coordinates are defined by the following three cross-ratios:

$$\begin{cases} k_0 = \{O''O'A,\ O''O'B;\ O''O'C,\ O''O'D\}, \\ k_1 = \{O'AO'',\ O'AB;\ O'AC,\ O'AD\}, \\ k_2 = \{AO''O',\ AO''B;\ AO''C,\ AO''D\}. \end{cases}$$

By intersecting the corresponding pencils of planes with the two image planes, we immediately obtain the values of k_0, k_1, k_2 as cross-ratios directly measurable in the two images:

$$\begin{cases} k_0 = \{e'a',\ e'b';\ e'c',\ e'd'\} = \{e''a'',\ e''b'';\ e''c'',\ e''d''\}, \\ k_1 = \{a'e',\ a'b';\ a'c',\ a'd'\}, \\ k_2 = \{a''e'',\ a''b'';\ a''c'',\ a''d''\}. \end{cases}$$

Figure 13.4 illustrates this method with data consisting of 46 point correspondences established between two images taken by weakly calibrated cameras. Figure 13.4(a) shows the input images and point matches. Figure 13.4(b) shows a view of the corresponding projective scene reconstruction, the raw projectives coordinates being used for rendering purposes. Since this form of display is not particularly enlightening, we also show in Figure 13.4(c) the reconstruction obtained by applying to the scene points the projective transformation mapping the three reference points (shown as small circles) and the camera centers onto their calibrated Euclidean positions. The true point positions are displayed as well for comparison.

13.2.2 Algebraic Motion Estimation

This section presents a purely algebraic approach to the problem of estimating the projective shape of a scene from binocular point correspondences, assuming once again that the stereo rig has been weakly calibrated. The perspective projection Eq. (2.15) introduced in chapter 2 extends naturally to the projective completion of $\mathbb{E}^3$ and maintains the same form in arbitrary projective frames for that space. Indeed, if we rewrite Eq. (2.15) as $z\boldsymbol{p} = \mathcal{M}\boldsymbol{P}$ in some Euclidean coordinate system (F), we obtain a similar equation $z\boldsymbol{p} = \mathcal{M}'\boldsymbol{P}'$ in a projective frame (F'), where $\boldsymbol{P}' = {}^{F'}P = {}^{F'}_{F}\mathcal{T}\boldsymbol{P}$, and $\mathcal{M}' = \mathcal{M}\,{}^{F'}_{F}\mathcal{T}^{-1}$.

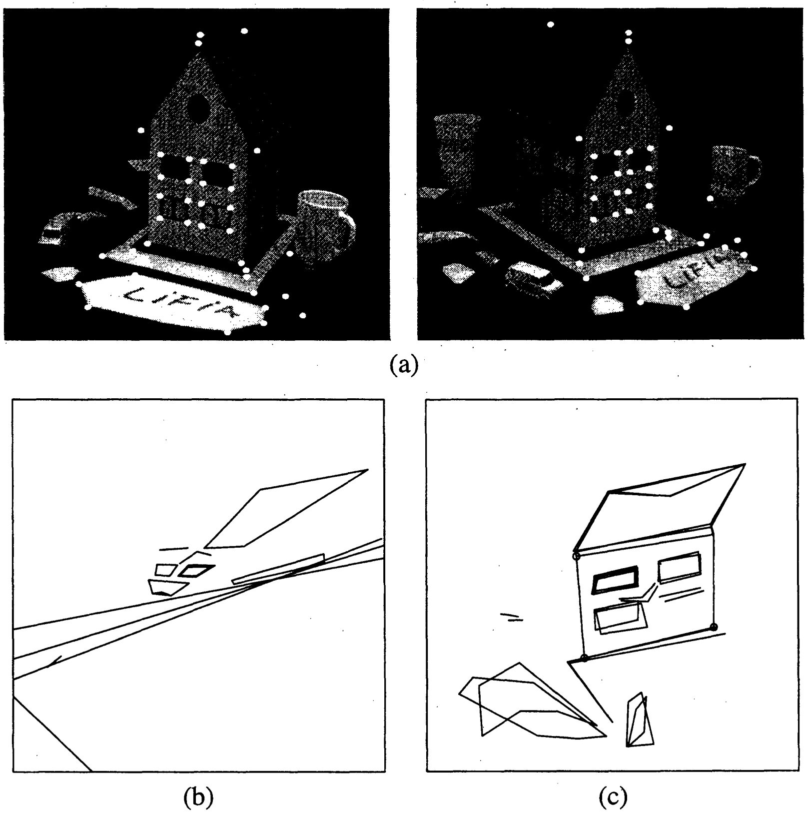


Figure 13.4 Geometric point reconstruction: (a) input data, (b) raw projective coordinates, (c) corrected projective coordinates. *Reprinted from "Relative Stereo and Motion Reconstruction," by J. Ponce, T.A. Cass, and D.H. Marimont, Tech. Report UIUC-BI-AI-RCV-93-07, Beckman Institute, Univ. of Illinois (1993). Data courtesy of Boubakeur Boufama and Roger Mohr.*

In particular, let us consider five points A_0, A_1, A_2, A_3, A_4 and choose them as a basis for $\tilde{\mathbb{E}}^3$, with A_4 playing the role of the unit point. We consider a camera observing these points, with projection matrix $\mathcal{M}$, and denote by a_0, a_1, a_2, a_3, a_4 the images of these points, choosing the points a_0 to a_3 as a projective basis of the image plane, a_3 being this time the unit point. We also denote by α, β, and γ the coordinates of a_4 in this basis.

Writing that $z_i \boldsymbol{a}_i = \mathcal{M}\boldsymbol{A}_i$ for $i = 0, 1, 2, 3, 4$ yields immediately

$$\mathcal{M} = \begin{pmatrix} z_0 & 0 & 0 & z_3 \\ 0 & z_1 & 0 & z_3 \\ 0 & 0 & z_2 & z_3 \end{pmatrix} \quad \text{and} \quad \begin{cases} z_4\alpha = z_0 + z_3, \\ z_4\beta = z_1 + z_3, \\ z_4\gamma = z_2 + z_3. \end{cases}$$

Since a perspective projection matrix is only defined up to scale, we can divide its coefficients by z_3, and defining $\lambda = z_4/z_3$ yields

$$\mathcal{M} = \begin{pmatrix} \lambda\alpha - 1 & 0 & 0 & 1 \\ 0 & \lambda\beta - 1 & 0 & 1 \\ 0 & 0 & \lambda\gamma - 1 & 1 \end{pmatrix}.$$

Let us now suppose we have a second image of the same scene, with projection matrix $\mathcal{M}'$ and image points $a'_0, a'_1, a'_2, a'_3, a'_4$. The same construction applies in this case, and we obtain

$$\mathcal{M}' = \begin{pmatrix} \lambda'\alpha' - 1 & 0 & 0 & 1 \\ 0 & \lambda'\beta' - 1 & 0 & 1 \\ 0 & 0 & \lambda'\gamma' - 1 & 1 \end{pmatrix}.$$

The stereo configuration of our two cameras is thus completely determined by the two parameters λ and λ'. The epipolar geometry of the rig can now be used to compute these parameters. Let us denote by C the optical center of the first camera and by e' the associated epipole in the image plane of the second camera, with coordinate vectors $\boldsymbol{C}$ and $\boldsymbol{e}'$ in the corresponding projective bases. We have $\mathcal{M}\boldsymbol{C} = \boldsymbol{0}$, and thus

$$\boldsymbol{C} = \left(\frac{1}{1-\lambda\alpha}, \frac{1}{1-\lambda\beta}, \frac{1}{1-\lambda\gamma}, 1\right)^T.$$

Substituting in the equation $\mathcal{M}'\boldsymbol{C} = \boldsymbol{e}'$ then yields

$$\boldsymbol{e}' = \left(1 - \frac{\lambda'\alpha' - 1}{\lambda\alpha - 1}, 1 - \frac{\lambda'\beta' - 1}{\lambda\beta - 1}, 1 - \frac{\lambda'\gamma' - 1}{\lambda\gamma - 1}\right)^T.$$

Now if μ' and ν' denote this time the *nonhomogeneous* coordinates of e' in the projective basis formed by the points a_i', we finally obtain

$$\begin{cases} \mu'(\lambda\gamma - \lambda'\gamma')(\lambda\alpha - 1) = (\lambda\alpha - \lambda'\alpha')(\lambda\gamma - 1), \\ \nu'(\lambda\gamma - \lambda'\gamma')(\lambda\beta - 1) = (\lambda\beta - \lambda'\beta')(\lambda\gamma - 1). \end{cases} \tag{13.5}$$

A system of two quadratic equations in two unknowns λ and λ' such as Eq. (13.5) admits in general four solutions, that can be thought of as the four intersections of the conic sections defined by the two equations in the (λ, λ') plane. Inspection of Eq. (13.5) reveals immediately that $(\lambda, \lambda') = (0, 0)$ and $(\lambda, \lambda') = (1/\gamma, 1/\gamma')$ are always solutions of these equations. It is easy (if a bit tedious) to show that the two remaining solutions are identical (geometrically, the two conics are tangent to each other at their point of intersection) and the corresponding values of the parameters λ and λ' are given by

$$\lambda = \frac{\operatorname{Det}\begin{pmatrix} \mu' & \alpha & \alpha' \\ \nu' & \beta & \beta' \\ 1 & \gamma & \gamma' \end{pmatrix}}{\operatorname{Det}\begin{pmatrix} \mu'\alpha & \alpha & \alpha' \\ \nu'\beta & \beta & \beta' \\ \gamma & \gamma & \gamma' \end{pmatrix}} \quad \text{and} \quad \lambda' = \frac{\operatorname{Det}\begin{pmatrix} \mu & \alpha & \alpha' \\ \nu & \beta & \beta' \\ 1 & \gamma & \gamma' \end{pmatrix}}{\operatorname{Det}\begin{pmatrix} \mu\alpha' & \alpha & \alpha' \\ \nu\beta' & \beta & \beta' \\ \gamma' & \gamma & \gamma' \end{pmatrix}}.$$

These values uniquely determine the projection matrices $\mathcal{M}$ and $\mathcal{M}'$. Note that taking into account the equations defining the second epipole would not add independent constraints because of the epipolar constraint that relates matching epipolar lines. Once the projection matrices are known, it is a simple matter to reconstruct the scene points.

13.3 PROJECTIVE MOTION ESTIMATION FROM MULTILINEAR CONSTRAINTS

The methods given in the previous two sections reconstruct the scene relative to five of its points, thus the quality of the reconstruction strongly depends on the accuracy of the localization of these points in the two images. In contrast, the approach presented in this section takes all points into account in a uniform manner and uses the multilinear constraints introduced in chapter 10 to reconstruct the camera motion in the form of the associated projection matrices.

13.3.1 Motion Estimation from Fundamental Matrices

Let us assume that the fundamental matrix $\mathcal{F}$ associated with two pictures has been estimated from binocular correspondences. As in the affine case, the projection matrices can in fact be estimated from a parameterization of $\mathcal{F}$ that exploits the inherent ambiguity of projective structure from motion. Since in the projective setting the scene structure and camera motion are only defined up to an arbitrary projective transformation, we can reduce the two matrices to canonical forms $\tilde{\mathcal{M}} = \mathcal{M}\mathcal{Q}$ and $\tilde{\mathcal{M}}' = \mathcal{M}'\mathcal{Q}'$ by postmultiplying them by an appropriate 4×4 matrix $\mathcal{Q}$. This time we take $\tilde{\mathcal{M}}'$ to be proportional to $(\text{Id} \quad \mathbf{0})$ and leave $\tilde{\mathcal{M}}$ in the general form $(\mathcal{A} \quad \boldsymbol{b})$. This reduction process determines 11 of the entries of $\mathcal{Q}$, and we refrain from using the 4 remaining degrees of freedom of $\mathcal{Q}$ to reduce $\tilde{\mathcal{M}}$ to a simpler form.

Let us now derive a new expression for the fundamental matrix using the canonical form of $\tilde{\mathcal{M}}'$. If $\tilde{\boldsymbol{P}} = (x, y, z, 1)^T$ denotes the homogeneous coordinate vector of the point P in the corresponding world coordinate system, we can write the projection equations associated with the two cameras as $z\boldsymbol{p} = (\mathcal{A} \quad \boldsymbol{b})\tilde{\boldsymbol{P}}$ and $z'\boldsymbol{p}' = (\text{Id} \quad \mathbf{0})\tilde{\boldsymbol{P}}$ or, equivalently,

$$z\boldsymbol{p} = \mathcal{A}(\text{Id} \quad \mathbf{0})\tilde{\boldsymbol{P}} + \boldsymbol{b} = z'\mathcal{A}\boldsymbol{p}' + \boldsymbol{b}.$$

It follows that $z\boldsymbol{b} \times \boldsymbol{p} = z'\boldsymbol{b} \times \mathcal{A}\boldsymbol{p}'$, and forming the dot product of this expression with $\boldsymbol{p}$ finally yields

$$\boldsymbol{p}^T \mathcal{F} \boldsymbol{p}' = 0 \quad \text{where} \quad \mathcal{F} = [\boldsymbol{b}_\times]\mathcal{A}.$$

Note the similarity with the expression for the essential matrix derived in chapter 10.

In particular, we have $\mathcal{F}^T\boldsymbol{b} = 0$, so (as could have been expected) $\boldsymbol{b}$ is the homogeneous coordinate vector of the first epipole in the corresponding image coordinate system. This new parameterization of the matrix $\mathcal{F}$ provides a simple method for computing the projection matrix $\tilde{\mathcal{M}}$. First note that, since the overall scale of $\tilde{\mathcal{M}}$ is irrelevant, we can always take $|\boldsymbol{b}| = 1$. This allows us to first compute $\boldsymbol{b}$ as the linear least-squares solution of $\mathcal{F}^T\boldsymbol{b} = 0$ with unit norm, then pick $\mathcal{A}_0 = -[\boldsymbol{b}_\times]\mathcal{F}$ as the value of $\mathcal{A}$. It is easy to show that, for any vector $\boldsymbol{a}$, $[\boldsymbol{a}_\times]^2 = \boldsymbol{a}\boldsymbol{a}^T - |\boldsymbol{a}|^2\text{Id}$; thus,

$$[\boldsymbol{b}_\times]\mathcal{A}_0 = -[\boldsymbol{b}_\times]^2\mathcal{F} = -\boldsymbol{b}\boldsymbol{b}^T\mathcal{F} + |\boldsymbol{b}|^2\mathcal{F} = \mathcal{F},$$

since $\mathcal{F}^T\boldsymbol{b} = \mathbf{0}$ and $|\boldsymbol{b}|^2 = 1$. This shows that $\tilde{\mathcal{M}} = \begin{pmatrix}\mathcal{A}_0 & \boldsymbol{b}\end{pmatrix}$ is *a* solution of our problem. As shown in the exercises, there is in fact a 4-parameter family of solutions whose general form is

$$\tilde{\mathcal{M}} = \begin{pmatrix}\mathcal{A} & \boldsymbol{b}\end{pmatrix} \quad \text{with} \quad \mathcal{A} = \lambda\mathcal{A}_0 + (\ \mu\boldsymbol{b} \mid \nu\boldsymbol{b} \mid \tau\boldsymbol{b}\ ).$$

The four parameters correspond, as could have been expected, to the remaining degrees of freedom of the projective transformation $\mathcal{Q}$. Once the matrix $\tilde{\mathcal{M}}$ is known, we can compute the position of any point P by solving in the least-squares sense the nonhomogeneous linear system of equations in z and z' defined by $z\boldsymbol{p} = z'\mathcal{A}\boldsymbol{p}' + \boldsymbol{b}$.

13.3.2 Motion Estimation from Trifocal Tensors

We now rewrite in a projective setting the trilinear constraints associated with the trifocal tensor first introduced in chapter 10. As in the previous section, we can postmultiply the projection matrices by an appropriate 4×4 matrix so they take the form

$$\tilde{\mathcal{M}}_1 = (\text{Id} \quad \mathbf{0}), \quad \tilde{\mathcal{M}}_2 = (\mathcal{A}_2 \quad \boldsymbol{b}_2), \quad \text{and} \quad \tilde{\mathcal{M}}_3 = (\mathcal{A}_3 \quad \boldsymbol{b}_3).$$

Under this transformation, $\boldsymbol{b}_2$ and $\boldsymbol{b}_3$ can still be interpreted as the homogeneous image coordinates of the epipoles e_{12} and e_{13}, and the trilinear constraints in Eqs. (10.14) and (10.15) still hold, with the trifocal tensor defined this time by the three matrices[2]

$$\mathcal{G}_1^i = \boldsymbol{b}_2 \boldsymbol{A}_3^{iT} - \boldsymbol{A}_2^i \boldsymbol{b}_3^T, \tag{13.6}$$

where $\boldsymbol{A}_2^i$ and $\boldsymbol{A}_3^i$ $(i = 1, 2, 3)$ denote the columns of $\mathcal{A}_2$ and $\mathcal{A}_3$.

Assuming that the trifocal tensor has been estimated from point or line correspondences as described in chapter 10, our goal in this section is to recover the projection matrices $\tilde{\mathcal{M}}_2$ and $\tilde{\mathcal{M}}_3$. Let us first observe that

$$(\boldsymbol{b}_2 \times \boldsymbol{A}_2^i)^T \mathcal{G}_1^i = \left[(\boldsymbol{b}_2 \times \boldsymbol{A}_2^i)^T \boldsymbol{b}_2\right] \boldsymbol{A}_3^{iT} - \left[(\boldsymbol{b}_2 \times \boldsymbol{A}_2^i)^T \boldsymbol{A}_2^i\right] \boldsymbol{b}_3^T = \boldsymbol{0},$$

and, likewise,

$$\mathcal{G}_1^i (\boldsymbol{b}_3 \times \boldsymbol{A}_3^i) = \left[\boldsymbol{A}_3^{iT} (\boldsymbol{b}_3 \times \boldsymbol{A}_3^i)\right] \boldsymbol{b}_2 - \left[\boldsymbol{b}_3^T (\boldsymbol{b}_3 \times \boldsymbol{A}_3^i)\right] \boldsymbol{A}_2^i = \boldsymbol{0}.$$

It follows that the matrix $\mathcal{G}_1^i$ is singular (a fact already mentioned in chapter 10) and the vectors $\boldsymbol{b}_2 \times \boldsymbol{A}_2^i$ and $\boldsymbol{b}_3 \times \boldsymbol{A}_3^i$ lie, respectively, in its left and right nullspaces. In turn, this means that, once the trifocal tensor is known, we can compute the epipole $\boldsymbol{b}_2$ (resp. $\boldsymbol{b}_3$) as the common normal to the left (resp. right) nullspaces of the matrices $\mathcal{G}_1^i$ $(i = 1, 2, 3)$.

Once the epipoles are known, writing Eq. (13.6) for $i = 1, 2, 3$ provides 27 homogeneous linear equations in the 18 unknown entries of the matrices $\mathcal{A}_j$ $(j = 2, 3)$. These equations can be solved up to scale using linear least squares. Alternatively, it is possible to estimate the matrices $\mathcal{A}_j$ directly from the trilinear constraints associated with pairs of matching points or lines by writing the trifocal tensor coefficients as functions of these matrices, which leads once again to a linear estimation process.

Once the projection matrices have been recovered, the projective structure of the scene can be recovered as well by using the perspective projection equations as linear constraints on the homogeneous coordinate vectors of the observed points and lines.

13.4 PROJECTIVE STRUCTURE AND MOTION FROM MULTIPLE IMAGES

Section 13.3 used the epipolar and trifocal constraints to reconstruct the camera motion and the corresponding scene structure from a pair or triple of images. Likewise, the quadrifocal tensor introduced in chapter 10 can in principle be used to estimate the projection matrices associated with four cameras and the corresponding projective scene structure. However, multilinear constraints do not provide a direct method for handling $m > 4$ views in a uniform manner. Instead, the structure and motion parameters estimated from pairs, triples, or quadruples of successive views must be stitched together iteratively. We now present an alternative where all images are taken into account at once in a nonlinear optimization scheme.

13.4.1 A Factorization Approach to Projective Structure from Motion

In this section, we present a factorization algorithm for motion analysis that generalizes the Tomasi–Kanade algorithm presented in chapter 12 to the projective case. Given m images of n

[2]Formally, postmultiplying the three projection matrices by $\mathcal{Q}$ has the same effect as taking the calibration matrix $\mathcal{K}_1$ equal to the identity in the equations defining the (uncalibrated) trifocal tensor in chapter 10. Note, however, that we do not assume here that the calibration parameters are known. Instead, we use an appropriate change of projective coordinates to simplify the form of the projection matrices.

points, we can rewrite Eq. (13.1) as

$$\mathcal{D} = \mathcal{M}\mathcal{P}, \tag{13.7}$$

where

$$\mathcal{D} \stackrel{\text{def}}{=} \begin{pmatrix} z_{11}\boldsymbol{p}_{11} & z_{12}\boldsymbol{p}_{12} & \dots & z_{1n}\boldsymbol{p}_{1n} \\ z_{21}\boldsymbol{p}_{21} & z_{22}\boldsymbol{p}_{22} & \dots & z_{2n}\boldsymbol{p}_{2n} \\ \dots & \dots & \dots & \dots \\ z_{m1}\boldsymbol{p}_{m1} & z_{m2}\boldsymbol{p}_{m2} & \dots & z_{mn}\boldsymbol{p}_{mn} \end{pmatrix}, \quad \mathcal{M} \stackrel{\text{def}}{=} \begin{pmatrix} \mathcal{M}_1 \\ \mathcal{M}_2 \\ \dots \\ \mathcal{M}_m \end{pmatrix} \text{ and } \mathcal{P} \stackrel{\text{def}}{=} (\boldsymbol{P}_1\ \boldsymbol{P}_2\ \dots\ \boldsymbol{P}_n).$$

As the product of $3m \times 4$ and $4 \times n$ matrices, the $3m \times n$ matrix $\mathcal{D}$ has (at most) rank 4; thus, if the projective depths z_{ij} were known, we could compute $\mathcal{M}$ and $\mathcal{P}$, just as in the affine case, by using singular value decomposition to factor $\mathcal{D}$. On the other hand, if $\mathcal{M}$ and $\mathcal{P}$ were known, we could read out the values of the projective depths z_{ij} from Eq. (13.7). This suggests an iterative scheme for estimating the unknowns z_{ij}, $\mathcal{M}$ and $\mathcal{P}$ by alternating steps where some of these unknowns are held constant while others are estimated.

We minimize the squared Frobenius norm of $\mathcal{D} - \mathcal{M}\mathcal{P}$—that is,

$$E \stackrel{\text{def}}{=} |\mathcal{D} - \mathcal{M}\mathcal{P}|^2 = \sum_{i,j} |z_{ij}\boldsymbol{p}_j - \mathcal{M}_i\boldsymbol{P}_j|^2$$

with respect to the unknowns $\mathcal{M}_i$, $\boldsymbol{P}_j$ and z_{ij}. Note that the minimization of E is ill-posed unless some constraints are imposed on the parameters $\mathcal{M}_i$, $\boldsymbol{P}_j$, and z_{ij}. Indeed, as mentioned earlier, these unknowns are not independent: The matrices $\mathcal{M}_i$ and the vectors $\boldsymbol{P}_j$ are only defined up to scale. If $\mathcal{M}_i$, $\boldsymbol{P}_j$, and z_{ij} are solutions of Eq. (13.1), so are $\alpha_i\mathcal{M}_i$, $\beta_j\boldsymbol{P}_j$, and $\alpha_i\beta_j z_{ij}$ for arbitrary values of the scalars α_i and β_j. In particular, Eq. (13.1) always admits the trivial solution $\mathcal{M}_i = 0, \boldsymbol{P}_j = 0, z_{ij} = 0$. In fact, this equation admits a much wider class of trivial, nonphysical solutions—for example, $z_{ij} = 0$, $\mathcal{M}_i = \mathcal{M}_0$, and $\boldsymbol{P}_j = \boldsymbol{P}_0$, where $\mathcal{M}_0$ is an arbitrary rank-3 3×4 matrix and $\boldsymbol{P}_0$ is a unit vector in its kernel. Here we impose the constraint that the columns $\boldsymbol{d}_j$ of the matrix $\mathcal{D}$ have unit norm, which eliminates these trivial solutions.

Let us assume that we are at some stage of the minimization process, fix the value of $\mathcal{M}$ to its current estimate and compute, for $j = 1, \dots, n$, the values of $z_j \stackrel{\text{def}}{=} (z_{1j}, \dots, z_{mj})^T$ and $\boldsymbol{P}_j$ that minimize

$$E_j \stackrel{\text{def}}{=} \sum_{i=1}^{m} |z_{ij}\boldsymbol{p}_j - \mathcal{M}_i\boldsymbol{P}_j|^2.$$

These values minimize E as well. Writing that the gradient of E_j with respect to the vector $\boldsymbol{P}_j$ should be zero at a minimum yields

$$0 = \frac{\partial E_j}{\partial \boldsymbol{P}_j} = 2\sum_{i=1}^{m} \mathcal{M}_i^T(z_{ij}\boldsymbol{p}_{ij} - \mathcal{M}_i\boldsymbol{P}_j),$$

or

$$\mathcal{M}^T\boldsymbol{d}_j = \mathcal{M}^T\mathcal{M}\boldsymbol{P}_j \iff \boldsymbol{P}_j = \mathcal{M}^\dagger\boldsymbol{d}_j,$$

where $\mathcal{M}^\dagger \stackrel{\text{def}}{=} (\mathcal{M}^T\mathcal{M})^{-1}\mathcal{M}^T$ is the pseudoinverse of $\mathcal{M}$. In turn, substituting this value in the definition of E_j yields $E_j = |(\text{Id} - \mathcal{M}\mathcal{M}^\dagger)\boldsymbol{d}_j|^2$.

Now $\mathcal{M}$ is a $3m \times 4$ matrix of rank 4 whose singular value decomposition $\mathcal{U}\mathcal{W}\mathcal{V}^T$ is formed by the product of a column-orthogonal $3m \times 4$ matrix $\mathcal{U}$, a 4×4 nonsingular diagonal matrix $\mathcal{W}$, and a 4×4 orthogonal matrix $\mathcal{V}^T$. The pseudoinverse of $\mathcal{M}$ is $\mathcal{M}^\dagger = \mathcal{V}\mathcal{W}^{-1}\mathcal{U}^T$; substituting this value in the expression of E_j and taking into account the fact that $|\boldsymbol{d}_j|^2 = 1$ immediately yields

$$E_j = \left|[\mathrm{Id} - \mathcal{U}\mathcal{U}^T]\boldsymbol{d}_j\right|^2 = 1 - |\mathcal{U}\boldsymbol{d}_j|^2.$$

In turn, this means that minimizing E_j with respect to z_j and $\boldsymbol{P}_j$ is equivalent to maximizing $|\mathcal{U}\boldsymbol{d}_j|^2$ under the constraint $|\boldsymbol{d}_j|^2 = 1$. Finally, observing that

$$\boldsymbol{d}_j = \mathcal{Q}_j z_j, \quad \text{where} \quad \mathcal{Q}_j \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{p}_{1j} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{p}_{2j} & \dots & \boldsymbol{0} \\ \dots & \dots & \dots & \dots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{p}_{mj} \end{pmatrix},$$

shows that minimizing E_j is equivalent to maximizing $|\mathcal{R}_j z_j|^2$ with respect to z_j under the constraint $|\mathcal{Q}_j z_j|^2 = 1$, where $\mathcal{R}_j \stackrel{\text{def}}{=} \mathcal{U}^T \mathcal{Q}_j$. This is a generalized eigenvalue problem, whose solution is the unit vector z_j corresponding to the largest scalar λ such that $\mathcal{R}_j^T \mathcal{R}_j z_j = \lambda \mathcal{Q}_j^T \mathcal{Q}_j z_j$.

The minimization step where the projective depths are held constant and $\mathcal{M}$ and $\mathcal{P}$ are updated is the same as in the Tomasi–Kanade approach to affine structure from motion. The overall process is summarized in Algorithm 13.1. The initial projective depth values are set to 1 or they can be computed as before from estimates of the epipolar geometry.

It is easy to show that the error E eventually converges to some value E^*. Indeed, let E_0 be the current error value at the beginning of each iteration; the first two steps of the algorithm do not change the vectors z_j, but minimize E with respect to the unknowns $\mathcal{M}$ and $\boldsymbol{P}_j$. If E_2 is the value of the error at the end of Step 2, we have $E_2 \leq E_0$. Now Step 3 does not change the matrix $\mathcal{M}$, but minimizes each error term E_j with respect to both the vectors z_j and $\boldsymbol{P}_j$. Therefore, the error E_3 at the end of this step is smaller than or equal to E_2. This shows that the error decreases in a monotone manner at each iteration. Since it is bounded below by zero, we conclude that the error converges to some value $E^* \geq 0$. The convergence of its error is not sufficient to guarantee the convergence of an optimization algorithm to a local minimum. However, a convergence proof for Algorithm 13.13.1, based on the *Global Convergence Theorem* (Luenberger, 1985) from numerical analysis and far too complex to be included here, can be found in Mahamud *et al.* (2001). Whether this local minimum turns out to be the global one depends, of course, on the choice of initial values chosen for the various unknown parameters. A

Algorithm 13.1: A Factorization Algorithm for Projective Shape from Motion.

1. Compute an initial estimate of the projective depths z_{ij}, with $i = 1, \dots, m$ and $j = 1, \dots, n$.

2. Normalize each column of the data matrix $\mathcal{D}$.

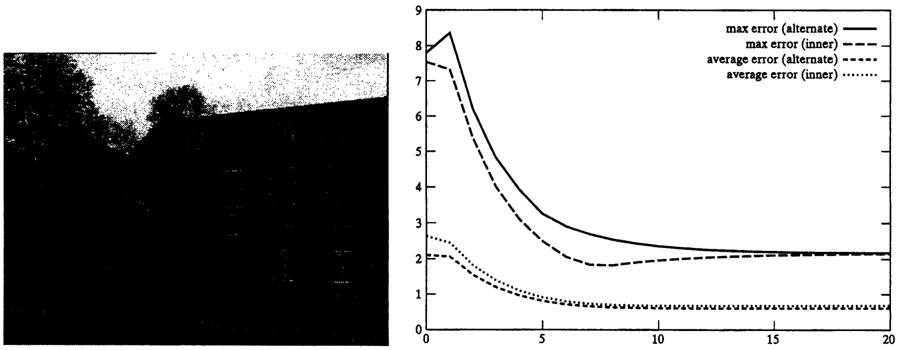
3. Repeat:

   (a) use singular value decomposition to compute the $2m \times 4$ matrix $\mathcal{M}$ and the $4 \times n$ matrix $\mathcal{P}$ that minimize $|\mathcal{D} - \mathcal{M}\mathcal{P}|^2$;

   (b) for $j = 1$ to n, compute the matrices $\mathcal{R}_j$ and $\mathcal{Q}_j$ and find the value of z_j that maximize $|\mathcal{R}_j z_j|^2$ under the constraint $|\mathcal{Q}_j z_j|^2 = 1$ as the solution of a generalized eigenvalue problem;

   (c) update the value of $\mathcal{D}$ accordingly;

until convergence.



(a) (b)

Figure 13.5 Iterative projective estimation of camera motion and scene structure: (a) a sample image in the sequence; (b) plot of the average and maximum reprojection error as a function of iteration number. Two experiments were conducted: In the first one (alternate), alternate images in the sequence are used as training and testing datasets; in the second experiment (inner), the first five and last five pictures were used as training set, and the remaining images were used for testing. In both cases, the average error falls below 1 pixel after 15 iterations. *Reprinted from "Iterative Projective Reconstruction from Multiple Views," by S. Mahamud and M. Hebert, Proc. IEEE Conference on Computer Vision and Pattern Recognition, (2000). © 2000 IEEE.*

possible choice, used in the experiments presented in Mahamud and Hebert (2000), is to initialize the projective depths z_{ij} to 1, which effectively amounts to starting with a weak-perspective projection model.

Figure 13.5(a) shows the first image in a sequence of 20 pictures of an outdoor scene. Thirty points were tracked manually across the sequence, with a localization error of ∓ 1 pixel. Figure 13.5(b) plots the evolution of the average and maximum errors between the observed and predicted image point positions when various subsets of the image sequence are used for training and testing.

13.4.2 Bundle Adjustment

Given initial estimates for the matrices $\mathcal{M}_i$ $(i = 1, \ldots, m)$ and vectors $\boldsymbol{P}_j$ $(j = 1, \ldots, n)$, we can refine these estimates by using nonlinear least squares to minimize the global error measure

$$E = \frac{1}{mn} \sum_{i,j} \left[\left(u_{ij} - \frac{\boldsymbol{m}_{i1} \cdot \boldsymbol{P}_j}{\boldsymbol{m}_{i3} \cdot \boldsymbol{P}_j} \right)^2 + \left(v_{ij} - \frac{\boldsymbol{m}_{i2} \cdot \boldsymbol{P}_j}{\boldsymbol{m}_{i3} \cdot \boldsymbol{P}_j} \right)^2 \right].$$

This is the method of *bundle adjustment*, whose name originates from the field of photogrammetry. Although it may be expensive, it offers the advantage of combining all measurements to minimize a physically significant error measure—namely, the mean-squared error between the actual image point positions and those predicted using the estimated scene structure and camera motion.

13.5 FROM PROJECTIVE TO EUCLIDEAN IMAGES

Although projective structure is useful by itself, in most cases it is the Euclidean structure of the scene that is the true object of interest. We saw in chapter 12 that the absolute scale of a scene cannot be recovered from weak-perspective or paraperspective images even when the intrinsic parameters of the corresponding cameras are known. The same ambiguity holds in the perspective case: Given a camera with known intrinsic parameters, we can take the calibration matrix to be the identity and write the perspective projection Eq. (2.15) in some Euclidean world coordinate system (W) as

$$p = \frac{1}{z}(\mathcal{R} \quad t)\begin{pmatrix} P \\ 1 \end{pmatrix} = \frac{1}{\lambda z}(\mathcal{R} \quad \lambda t)\begin{pmatrix} \lambda P \\ 1 \end{pmatrix}$$

for any nonzero scale factor λ. This ambiguity is not surprising given the fact that t is defined in Eq. (2.15) as the position of the origin of (W) relative to the camera: Moving both the scene and the camera observing it away from (or toward) this point at constant speed alters the apparent depth of the scene, but does not change its image. Adding more cameras does not help, and the best we can hope for is to estimate the Euclidean shape of the scene, defined, as in chapter 2, up to an arbitrary similarity transformation.

Let us assume from now on that one of the techniques presented in Section 13.4 has been used to estimate the projection matrices $\mathcal{M}_i$ $(i = 1, \dots, m)$ and the point positions P_j $(j = 1, \dots, n)$ from m images of these points. We know that any other reconstruction *and in particular a Euclidean one* is separated from this one by a projective transformation. In other words, if $\hat{\mathcal{M}}_i$ and $\hat{P}_j$ denote the shape and motion parameters measured in some Euclidean coordinate system, there must exist a 4×4 matrix $\mathcal{Q}$ such that $\hat{\mathcal{M}}_i = \mathcal{M}_i \mathcal{Q}$ and $\hat{P}_j = \mathcal{Q}^{-1} P_j$. The rest of this section presents a method for computing the *Euclidean upgrade* matrix $\mathcal{Q}$ and thus recovering the Euclidean shape and motion from the projective ones when (some of) the intrinsic parameters of the camera are known.

Let us first note that, since the individual matrices $\mathcal{M}_i$ are only defined up to scale, so are the matrices $\hat{\mathcal{M}}_i$ that can be written (in the most general case where some of the intrinsic parameters are unknown) as

$$\hat{\mathcal{M}}_i = \rho_i \mathcal{K}_i (\mathcal{R}_i \quad t_i),$$

where ρ_i accounts for the unknown scale of $\mathcal{M}_i$, and $\mathcal{K}_i$ is a calibration matrix as defined by Eq. (2.13). In particular, if we write the Euclidean upgrade matrix as $\mathcal{Q} = (\mathcal{Q}_3 \quad q_4)$, where $\mathcal{Q}_3$ is a 4×3 matrix and q_4 is a vector in $\mathbb{R}^4$, we obtain immediately

$$\mathcal{M}_i \mathcal{Q}_3 = \rho_i \mathcal{K}_i \mathcal{R}_i. \tag{13.8}$$

Using this equation, it is a simple matter to adapt the affine methods introduced in chapter 12 to the projective setting when the intrinsic parameters of all cameras are known so the matrices $\mathcal{K}_i$ can be taken equal to the identity: According to Eq. (13.8), the 3×3 matrices $\mathcal{M}_i \mathcal{Q}_3$ are in this case scaled rotation matrices. Writing that their rows m_{ij}^T $(j = 1, 2, 3)$ are perpendicular to each other and have the same norm yields

$$\begin{cases} m_{i1}^T \mathcal{Q}_3 \mathcal{Q}_3^T m_{i2} = 0, \\ m_{i2}^T \mathcal{Q}_3 \mathcal{Q}_3^T m_{i3} = 0, \\ m_{i3}^T \mathcal{Q}_3 \mathcal{Q}_3^T m_{i1} = 0, \\ m_{i1}^T \mathcal{Q}_3 \mathcal{Q}_3^T m_{i1} - m_{i2}^T \mathcal{Q}_3 \mathcal{Q}_3^T m_{i2} = 0, \\ m_{i2}^T \mathcal{Q}_3 \mathcal{Q}_3^T m_{i2} - m_{i3}^T \mathcal{Q}_3 \mathcal{Q}_3^T m_{i3} = 0. \end{cases} \tag{13.9}$$

The upgrade matrix $\mathcal{Q}$ is of course only defined up to an arbitrary similarity. To determine it uniquely, we can assume that the world coordinate system and the first camera's frame coincide. Given m images, we obtain 12 linear equations and $5(m - 1)$ quadratic ones in the coefficients of $\mathcal{Q}$. These equations can be solved using nonlinear least squares.

Alternatively, the constraints in Eq. (13.9) are linear in the coefficients of the symmetric matrix $\mathcal{A} \stackrel{\text{def}}{=} \mathcal{Q}_3\mathcal{Q}_3^T$, allowing its estimation from at least two images via linear least squares. Note that $\mathcal{A}$ has rank 3—a constraint not enforced by our construction. To recover $\mathcal{Q}_3$, let us also note that, since $\mathcal{A}$ is symmetric, it can be diagonalized in an orthonormal basis as $\mathcal{A} = \mathcal{U}\mathcal{D}\mathcal{U}^T$, where $\mathcal{D}$ is the diagonal matrix formed by the eigenvalues of $\mathcal{A}$ and $\mathcal{U}$ is the orthogonal matrix formed by its eigenvectors. In the absence of noise, $\mathcal{A}$ is positive semidefinite with three positive and one zero eigenvalues, and $\mathcal{Q}_3$ can be computed as $\mathcal{U}_3\sqrt{\mathcal{D}_3}$, where $\mathcal{U}_3$ is the matrix formed by the columns of $\mathcal{U}$ associated with the positive eigenvalues of $\mathcal{A}$, and $\mathcal{D}_3$ is the corresponding submatrix of $\mathcal{D}$. Because of noise, however, $\mathcal{A}$ usually has maximal rank, and its smallest eigenvalue may even be negative. As shown in Ponce (2000), if we take this time $\mathcal{U}_3$ and $\mathcal{D}_3$ to be the submatrices of $\mathcal{U}$ and $\mathcal{D}$ associated with the three largest (positive) eigenvalues of $\mathcal{A}$, then $\mathcal{U}_3\mathcal{D}_3\mathcal{U}_3^T$ provides the best positive semidefinite rank-3 approximation of $\mathcal{A}$ in the sense of the Frobenius norm,[3] and we can take as before $\mathcal{Q}_3 = \mathcal{U}_3\sqrt{\mathcal{D}_3}$. At this point, the last column vector $\boldsymbol{q}_4$ of $\mathcal{Q}$ can be determined by (arbitrarily) picking the origin of the frame attached to the first camera as the origin of the world coordinate system.

This method can easily be adapted to the case where only some of the intrinsic camera parameters are known: Using the fact that $\mathcal{R}_i$ is an orthogonal matrix allows us to write

$$\mathcal{M}_i\mathcal{A}\mathcal{M}_i^T = \rho_i^2\mathcal{K}_i\mathcal{K}_i^T. \tag{13.10}$$

Thus, every image provides a set of constraints between the entries of $\mathcal{K}_i$ and $\mathcal{A}$. Assuming, for example, that the center of the image is known for each camera, we can take $u_0 = v_0 = 0$ and write the square of the matrix $\mathcal{K}_i$ as

$$\mathcal{K}_i\mathcal{K}_i^T = \begin{pmatrix} \alpha_i^2\dfrac{1}{\sin^2\theta_i} & -\alpha_i\beta_i\dfrac{\cos\theta_i}{\sin^2\theta_i} & 0 \\ -\alpha_i\beta_i\dfrac{\cos\theta_i}{\sin^2\theta_i} & \beta_i^2\dfrac{1}{\sin^2\theta_i} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In particular, the part of Eq. (13.10) corresponding to the zero entries of $\mathcal{K}_i\mathcal{K}_i^T$ provides two independent linear equations in the 10 coefficients of the 4×4 symmetric matrix $\mathcal{A}$. With $m \geq 5$ images, these parameters can be estimated via linear least squares. Once $\mathcal{A}$ is known, $\mathcal{Q}$ can be estimated as before. Figure 13.6 shows a texture-mapped picture of the 3D model of a castle obtained by a variant of this method (Pollefeys *et al.* (1999)).

13.6 NOTES

The short introduction to projective geometry given at the beginning of this chapter focuses on the analytical side of things. See, for example, Todd (1946), Berger (1987), and Samuel (1988) for thorough introductions to analytical projective geometry, and Coxeter (1974) for an axiomatic presentation. Projective structure from motion is covered in detail in the books of Hartley and Zisserman (2000) and Faugeras, Luong, and Papadopoulo (2001).

As mentioned by Faugeras (1993), the problem of calculating the epipoles and the epipolar transformations compatible with seven point correspondences was first posed by Chasles (1855) and solved by Hesse (1863). The problem of estimating the epipolar geometry from five point correspondences for internally calibrated cameras was solved by Kruppa (1913). An excellent

[3]Note the obvious similarity between this result and Theorem 4.

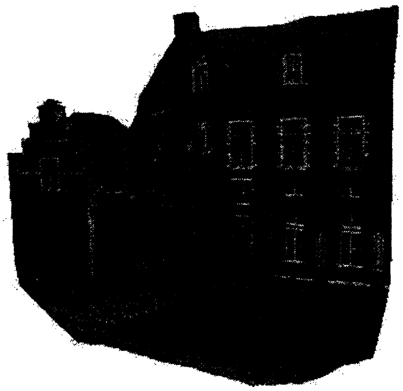


Figure 13.6 A synthetic texture-mapped image of a castle constructed via projective motion analysis followed by a Euclidean upgrade. The principal point is assumed to be known. *Reprinted from "Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences," by M. Pollefeys, PhD Thesis, Katholieke Universiteit, Leuven, (1999).*

modern account of Hesse's and Kruppa's techniques can be found in Faugeras and Maybank (1990), where the *absolute conic*, an imaginary conic section invariant through similarities, is used to derive two tangency constraints that make up for the missing point correspondences. These methods are of course mostly of theoretical interest since their reliance on a minimal number of correspondences limits their ability to deal with noise. The weak-calibration methods of Luong *et al.* (1993, 1996) and Hartley (1995) described in chapter 10 provide reliable and accurate alternatives.

Faugeras (1992) and Hartley *et al.* (1992) introduced independently the idea of using a pair of uncalibrated cameras to recover the projective structure of a scene. Other notable work in this area includes, for example, Mohr *et al.* (1992) and Shashua (1993). Section 13.2.2 presents Faugeras' original method, and its geometric variant presented in Section 13.2.1 is taken from Ponce *et al.* (1993). The two- and three-view motion analysis techniques also presented in this chapter are variants of the methods proposed by Hartley (1992, 1994*b*, 1997) and Beardsley *et al.* (1997). When the cameras are calibrated, it is also possible, as shown in the exercises and (Longuet–Higgins, 1981), to recover (up to a two-fold ambiguity) the similitude associated with the corresponding *essential* matrix. An iterative algorithm for perspective motion and structure recovery using calibrated cameras is given in Christy and Horaud (1996). The extension of factorization approaches to structure and motion recovery was first proposed by Sturm and Triggs (1996). The variant presented in Section 13.4.1 is due to Mahamud and Hebert (2000) and has the advantage of being provably convergent (Mahamud *et al.*, 2001). Algorithms for stitching together pairs, triples or quadruples of successive views can be found in Beardsley *et al.* (1997) and Pollefeys *et al.* (1999) for example.

The problem of computing Euclidean upgrades of projective reconstructions when some of the intrinsic parameters are known has been addressed by a number of authors (e.g., Heyden and Åström, 1996, Triggs, 1997, Pollefeys, 1999). The matrix $\mathcal{A} = \mathcal{Q}_3\mathcal{Q}_3^T$ introduced in Section 13.5 can be interpreted geometrically as the projective representation of the dual of the absolute conic, the *absolute dual quadric* (Triggs, 1997). Like the absolute conic, this quadric surface is invariant through similarities, and the (dual) conic section associated with $\mathcal{K}_i\mathcal{K}_i^T$ is simply the projection of this quadric surface into the corresponding image. Self-calibration is the process of computing the intrinsic parameters of a camera from point correspondences with unknown Euclidean positions. Work in this area was pioneered by Faugeras and Maybank (1992) for cameras with fixed intrinsic parameters. A number of reliable self-calibration methods are now available (Hartley, 1994*a*, Fitzgibbon and Zisserman, 1998, Pollefeys *et al.*, 1999), and they can also be used to upgrade projective reconstructions to Euclidean ones. The problem of computing Euclidean upgrades of projective reconstructions under minimal camera constraints such as zero skew is addressed in Heyden and Åström (1998, 1999), Pollefeys *et al.* (1999), and Ponce (2000).

PROBLEMS

13.1. Use a simple counting argument to determine the minimum number of point correspondences required to solve the projective structure-from-motion problem in the trinocular case.

13.2. Show that the change of coordinates between two projective frames (A) and (B) can be represented by Eq. (13.2).

13.3. Show that any two distinct lines in a projective plane intersect in exactly one point and that two parallel lines Δ and Δ' in an affine plane intersect at the point at infinity associated with their common direction $\boldsymbol{v}$ in the projective completion of this plane.

Hint: Use J_A to embed the affine plane in its projective closure, and write the vector of $\Pi \times \mathbb{R}$ associated with any point in $J_A(\Delta)$ (resp. $J_A(\Delta')$) as a linear combination of the vectors $(\overrightarrow{AB}, 1)$ and $(\overrightarrow{AB} + \boldsymbol{v}, 1)$ (resp. $(\overrightarrow{AB'}, 1)$ and $(\overrightarrow{AB'} + \boldsymbol{v}, 1)$), where B and B' are arbitrary points on Δ and Δ'.

13.4. Show that a perspective projection between two planes of $\mathbb{P}^3$ is a projective transformation.

13.5. Given an affine space X and an affine frame $(A_0, \ldots, A_n)$ for that space, what is the projective basis of $\tilde{X}$ associated with the vectors $\boldsymbol{e}_i \stackrel{\text{def}}{=} (\overrightarrow{A_0 A_i}, 0)$ $(i = 1, \ldots, n)$ and the vector $\boldsymbol{e}_{n+1} = (\mathbf{0}, 1)$? Are the points $J_{A_0}(A_i)$ part of that basis?

13.6. In this exercise, you will show that the cross-ratio of four collinear points A, B, C, and D is equal to

$$\{A, B; C, D\} = \frac{\sin(\alpha + \beta)\sin(\beta + \gamma)}{\sin(\alpha + \beta + \gamma)\sin\beta},$$

where the angles α, β, and γ are defined as in Figure 13.2.

(a) Show that the area of a triangle PQR is

$$A(P, Q, R) = \frac{1}{2} PQ \times RH = \frac{1}{2} PQ \times PR \sin\theta,$$

where PQ denotes the distance between the two points P and Q, H is the projection of R onto the line passing through P and Q, and θ is the angle between the lines joining the point P to the points Q and R.

(b) Define the ratio of three collinear points A, B, C as

$$R(A, B, C) = \frac{\overline{AB}}{\overline{BC}}$$

for some orientation of the line supporting the three points. Show that

$$R(A, B, C) = A(A, B, O)/A(B, C, O),$$

where O is some point not lying on this line.

(c) Conclude that the cross-ratio $\{A, B; C, D\}$ is indeed given by the formula above.

13.7. Show that the homography between two epipolar pencils of lines can be written as

$$\tau \to \tau' = \frac{a\tau + b}{c\tau + d},$$

where τ and τ' are the slopes of the lines.

13.8. Here we revisit the three-point reconstruction problem in the context of the *homogeneous* coordinates of the point D in the projective basis formed by the tetrahedron (A, B, C, O') and the unit point O''. Note that the ordering of the reference points, and thus the ordering of the coordinates, is different from the one used earlier: This new choice is, like the previous one, made to facilitate the reconstruction.

We denote the (unknown) coordinates of the point D by (x, y, z, w), equip the first (resp. second) image plane with the triangle of reference a', b', c' (resp. a'', b'', c'') and the unit point e' (resp. e''), and denote by (x', y', z') (resp. (x'', y'', z'')) the coordinates of the point d' (resp. d'').

Hint: Drawing a diagram similar to Figure 13.3 helps.

(a) What are the homogeneous projective coordinates of the points D', D'', and E where the lines $O'D$, $O''D$, and $O'O''$ intersect the plane of the triangle?

(b) Write the coordinates of D as a function of the coordinates of O' and D' (resp. O'' and D'') and some unknown parameters.

Hint: Use the fact that the points D, O', and D' are collinear.

(c) Give a method for computing these unknown parameters and the coordinates of D.

13.9. Show that if $\tilde{\mathcal{M}} = (\mathcal{A} \quad \boldsymbol{b})$ and $\tilde{\mathcal{M}}' = (\mathrm{Id} \quad \mathbf{0})$ are two projection matrices, and if $\mathcal{F}$ denotes the corresponding fundamental matrix, then $[\boldsymbol{b}_\times]\mathcal{A}$ is proportional to $\mathcal{F}$ whenever $\mathcal{F}^T\boldsymbol{b} = 0$ and

$$\mathcal{A} = -\lambda[\boldsymbol{b}_\times]\mathcal{F} + \left(\mu\boldsymbol{b} \mid \nu\boldsymbol{b} \mid \tau\boldsymbol{b} \right).$$

13.10. We derive in this exercise a method for computing a minimal parameterization of the fundamental matrix and estimating the corresponding projection matrices. This is similar in spirit to the technique presented in Section 12.2.2 of Chapter 12 in the affine case.

(a) Show that two projection matrices $\mathcal{M}$ and $\mathcal{M}'$ can always be reduced to the following canonical forms by an appropriate projective transformation:

$$\tilde{\mathcal{M}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \tilde{\mathcal{M}}' = \begin{pmatrix} \boldsymbol{a}_1^T & b_1 \\ \boldsymbol{a}_2^T & b_2 \\ \mathbf{0}^T & 1 \end{pmatrix}.$$

Note: For simplicity, you can assume that all the matrices involved in your solution are nonsingular.

(b) Note that applying this transformation to the projection matrices amounts to applying the inverse transformation to every scene point P. Let us denote by $\tilde{\boldsymbol{P}} = (x, y, z)^T$ the position of the transformed point $\tilde{P}$ in the world coordinate system and by $\boldsymbol{p} = (u, v, 1)^T$ and $\boldsymbol{p}' = (u', v', 1)^T$ the homogeneous coordinate vectors of its images. Show that

$$(u' - b_1)(\boldsymbol{a}_2 \cdot \boldsymbol{p}) = (v' - b_2)(\boldsymbol{a}_1 \cdot \boldsymbol{p}).$$

(c) Derive from this equation an eight-parameter parameterization of the fundamental matrix, and use the fact that $\mathcal{F}$ is only defined up to a scale factor to construct a minimal seven-parameter parameterization.

(d) Use this parameterization to derive an algorithm for estimating $\mathcal{F}$ from at least seven point correspondences and for estimating the projective shape of the scene.

13.11. Here we address the problem of recovering the rotation $\mathcal{R}$ and translation $\boldsymbol{t}$ associated with an essential matrix $\mathcal{E} = [\boldsymbol{t}_\times]\mathcal{R}$ (this exercise is courtesy of Andrew Zisserman). The translation part is easy since $\boldsymbol{t}$ can be recovered (up to scale since we know that the structure of a scene can only be determined up to a similitude) as the unit vector satisfying $\mathcal{E}^T\boldsymbol{t}$.

(a) Show that the SVD of the essential matrix can be written as

$$\mathcal{E} = \mathcal{U}\ \mathrm{diag}(1, 1, 0)\mathcal{V}^\top,$$

and conclude that $\boldsymbol{t}$ is the third column vector of $\mathcal{U}$.

(b) Show that the two matrices

$$\mathcal{R}_1 = \mathcal{U}\mathcal{W}\mathcal{V}^\top \quad \mathcal{R}_2 = \mathcal{U}\mathcal{W}^\top\mathcal{V}^\top$$

satisfy $\mathcal{E} = [\boldsymbol{t}]_\times \mathcal{R}$, where

$$\mathcal{W} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Programming Assignments

13.12. Implement the geometric approach to projective scene estimation introduced in Section 13.2.1.

13.13. Implement the algebraic approach to projective scene estimation introduced in Section 13.2.2.

13.14. Implement the factorization approach to projective scene estimation introduced in Section 13.4.1.