2 Theory of close range photogrammetry

M.A.R. Cooper with S. Robson

Photogrammetry is a technique for obtaining information about the position, size and shape of an object by measuring images of it instead of by measuring it directly. The term 'close range photogrammetry' is used to describe the technique when the extent of the object to be measured is less than about 100 metres and cameras are positioned close to it. Other characteristics have come to be associated with close range photogrammetry which make it different from aerial mapping. Images are obtained from camera positions all around (and sometimes inside) the object (Figure 2.1). Camera axes are parallel only in special cases; usually they are highly convergent, pointing generally towards the middle of the object. The co-ordinates of points on the surface of an object are often required to be of high homogeneous accuracy throughout the object space. Based on an understanding of the physical processes of imaging and measurement, mathematical models are devised which form the basis of numerical methods to produce three dimensional co-ordinates of discrete points on the object. These co-ordinates are usually estimated by least squares, sometimes with thousands of degrees of freedom. One of the results of the increasing use of digital cameras for photogrammetry is the transfer of machine vision algorithms and concepts into photogrammetric processes. Image features can now be automatically identified, matched and transformed into three dimensional features in object space.

The results of close range photogrammetry must generally be made available very quickly after acquisition of the images so that they can be used for further processing related to the measured object and its function. Derived co-ordinates might be used for comparing the measured object with its designed size and shape in a test of conformance. Or they may be compared with a previous set of co-ordinates to detect deformation of the object. They are sometimes processed further using computer graphics, for example to produce a three dimensional CAD model of the object (Figure 2.1) and in a few cases drawn and dimensioned plans, elevations or sections may be required. Another significant characteristic of close range photogrammetry is the great diversity of measurement problems that can be solved using the technique. They call for a range of cameras, imaging media, configurations, photogrammetric procedures, methods of analysis and form of results to be considered and for specific instrumentation and techniques to be selected and used in each particular case to produce results that meet specifications (*see* Chapter 9).

The approach taken in this Chapter to give a theoretical basis for close range photogrammetry is now summarized. The reader is assumed to be familiar with basic statis-

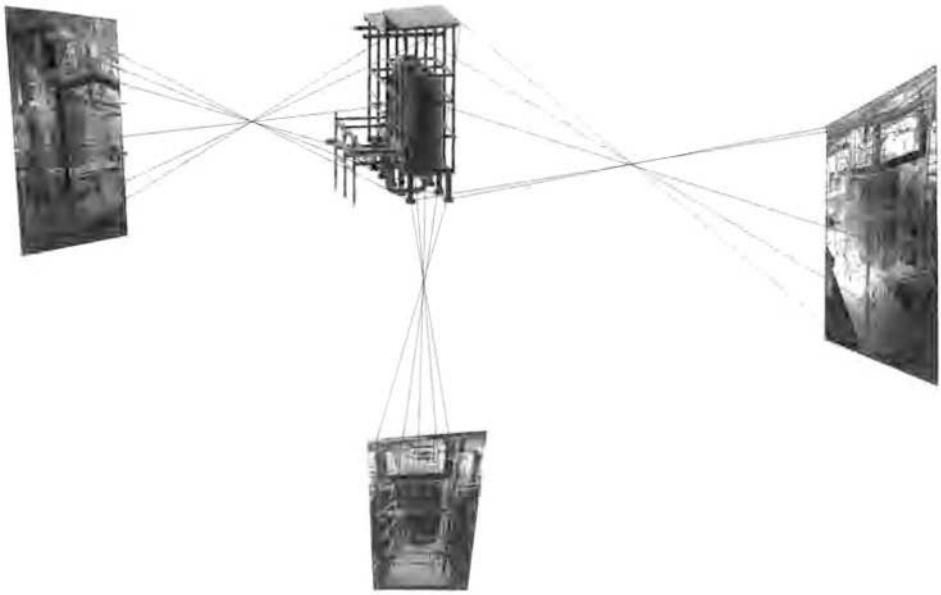

Figure 2.1 *Close range, multistation, convergent camera configuration. Only three of many cameras are illustrated.*

tics and matrix and vector algebra. Cartesian co-ordinates are convenient for defining positions of points, so co-ordinate transformations are discussed (section 2.1). A geometrical description of the complicated physical processes by which an image of a feature on the object is produced by a camera is described (section 2.2) followed by the geometrical relationships between two cameras and an object (section 2.3). The geometry of all cameras and an object is used (section 2.4) for deriving equations that are the basis of least squares estimation as a process for transforming measurements in the images into spatial information about the object. Theoretical and practical aspects of least squares estimation are discussed (section 2.5), and the Chapter ends with a consideration of indicators of the quality of data, whether measured or derived by least squares estimation (section 2.6). The general nature of the chapter means that few specific references are given. However, a bibliography is included so that the reader can find more information about the relevant principal subjects than has been possible to include here.

2.1 Co-ordinates and co-ordinate transformations

In photogrammetry the position of a point in space is commonly defined by a three dimensional cartesian co-ordinate system, the origin, scale and orientation of which can be arbitrarily defined. It is often necessary to convert between co-ordinates in systems having different origins, orientations and possibly scales. For example, in photogrammetry of a small industrial component it may be convenient initially to define co-ordinates of points on it with reference to a co-ordinate datum related to features on the

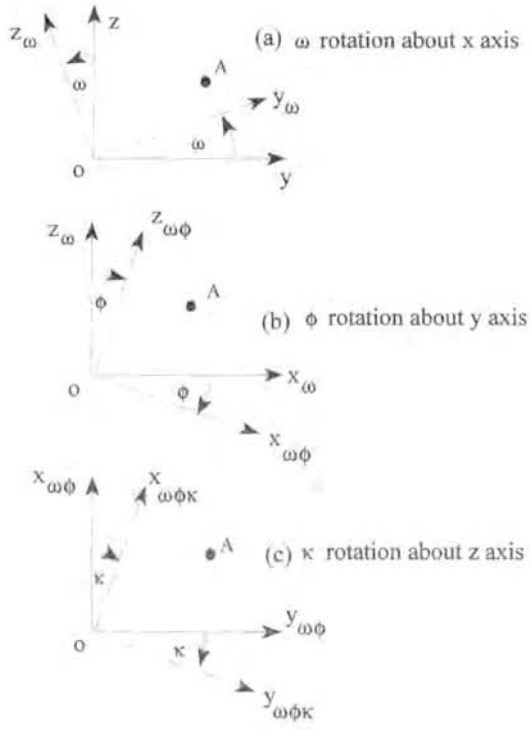

Figure 2.2 *Sequential rotations of axes in three dimensional space.*

object itself. If it is later relocated into a larger component, it may then be necessary to define co-ordinates of points on it relative to a new datum.

Co-ordinate transformations may be divided into three parts: scale change, translation and rotation. A scale change λ along each of the three axes may be represented by the vector equation $\mathbf{x} = \lambda \mathbf{X}$, where $\mathbf{X} = [X\ Y\ Z]^t$ is the position vector of a point in the primary co-ordinate system, and $\mathbf{x} = [x\ y\ z]^t$ is the position vector of the point in the secondary (scaled) co-ordinate system. A translation of axes may be represented by the vector equation $\mathbf{x} = \mathbf{X} - \mathbf{X}_o$, where $\mathbf{X} = [X\ Y\ Z]^t$, is the position vector of a point in the primary co-ordinate system, $\mathbf{X}_o = [X_o\ Y_o\ Z_o]^t$, is the position vector of the origin of the secondary co-ordinate system relative to the primary, and $\mathbf{x} = [x\ y\ z]^t$, is the position vector of the point in the secondary co-ordinate system.

2.1.1 Rotation matrices

Co-ordinate transformations arising from rotations of orthogonal axes in three dimensional space may be expressed as the resultant of three independent sequential transformations. Figure 2.2(*a*) shows a point A with co-ordinates (x, y, z) relative to the (xyz) axes. If a rotation ω is made clockwise about the (positive) x-axis, the position vector of A in the rotated system $(x_\omega y_\omega z_\omega)$ is given by the vector equation $[x_\omega\ y_\omega\ z_\omega]^t = \mathbf{R}_\omega [x\ y\ z]^t$ where

$$\mathbf{R}_{\omega} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{bmatrix}. \tag{2.1}$$

If the axes are now given a rotation φ clockwise about the y_{ω} axis (Figure 2.2*b*) the co-ordinates of A in the $(x_{\omega\varphi} y_{\omega\varphi} z_{\omega\varphi})$ system will be $[x_{\omega\varphi}\ y_{\omega\varphi}\ z_{\omega\varphi}]^t = \mathbf{R}_{\varphi}\,\mathbf{R}_{\omega}\,[x\ y\ z]^t$ where

$$\mathbf{R}_{\varphi} = \begin{bmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{bmatrix}. \tag{2.2}$$

Finally, after a rotation κ clockwise about the $z_{\omega\varphi}$ axis (Figure 2.2*c*), the co-ordinates of A in the $(x_{\omega\varphi\kappa} y_{\omega\varphi\kappa} z_{\omega\varphi\kappa})$ system will be $[x_{\omega\varphi\kappa}\ y_{\omega\varphi\kappa}\ z_{\omega\varphi\kappa}]^t = \mathbf{R}_{\kappa}\,\mathbf{R}_{\varphi}\,\mathbf{R}_{\omega}\,[x\ y\ z]^t$ where

$$\mathbf{R}_{\kappa} = \begin{bmatrix} \cos\kappa & \sin\kappa & 0 \\ -\sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.3}$$

The matrix product $\mathbf{R}_{\kappa}\,\mathbf{R}_{\varphi}\,\mathbf{R}_{\omega}$ corresponding to primary rotation ω, secondary rotation φ and tertiary rotation κ can be denoted by $\mathbf{R}_{\omega\varphi\kappa}$ and is, in full:

$$\mathbf{R}_{\omega\varphi\kappa} = \mathbf{R}_{\kappa}\mathbf{R}_{\varphi}\mathbf{R}_{\omega} = \begin{bmatrix} \cos\varphi\cos\kappa & \sin\omega\sin\varphi\cos\kappa + \cos\omega\sin\kappa & -\cos\omega\sin\varphi\cos\kappa + \sin\omega\sin\kappa \\ -\cos\varphi\sin\kappa & -\sin\omega\sin\varphi\sin\kappa + \cos\omega\cos\kappa & \cos\omega\sin\varphi\cos\kappa + \sin\omega\cos\kappa \\ \sin\varphi & -\sin\omega\cos\varphi & \cos\omega\cos\varphi \end{bmatrix} \tag{2.4}$$

or

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}. \tag{2.5}$$

If the elements of $\mathbf{R}$ are known, the rotations ω, φ and κ can be found from

$$\sin\varphi = r_{31}, \qquad \tan\omega = -r_{32}/r_{33}, \text{ and } \tan\kappa = -r_{21}/r_{11}$$

but in each case, two values of each angle between 0 and 2π radians are possible. If the first value of φ from $\sin\varphi = r_{31}$ (say φ_1) is substituted in r_{32} it will give two values for ω. If φ_1 is substituted in r_{33} two more values of ω will be obtained. Of these four values of ω, two will be the same and this value (say ω_1) is the value of ω corresponding to φ_1; ω_2 can be found from φ_2 in the same way. Similarly, κ_1 and κ_2 can be evaluated from φ_1 and φ_2 by substitutions in r_{11} and r_{21}. If the values and sequence of three rotation angles are

given, the values of the nine elements of the corresponding rotation matrix are uniquely defined. However given the values of the nine elements of a rotation matrix having a specific sequence of rotations, two sets of the three rotation angles can be evaluated.

The rotation matrix **R** is orthogonal, so $\mathbf{R}^{-1} = \mathbf{R}^{t}$. The individual rotation matrices are not commutable in multiplication, so if the rotations are performed in a different order from that given above, the algebraic form in (2.4) will be different. For example if the first rotation is φ, the second ω and the last κ, the full rotation matrix will be

$$\mathbf{R}_{\varphi\omega\kappa} = \mathbf{R}_{\kappa}\mathbf{R}_{\omega}\mathbf{R}_{\varphi}.$$

Primary co-ordinates (X, Y, Z) of a point are transformed to secondary co-ordinates (x, y, z) by the rotation matrix **R** as follows:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \tag{2.6a}$$

The rotation matrix **R** is orthogonal, so the reverse transformation is:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{2.6b}$$

Other forms of 3 × 3 orthogonal matrices with elements expressed as functions of three independent parameters may be used for **R**. Elements of the *Rodrigues matrix* for example are algebraic functions of three independent parameters a, b and c:

$$R = \Delta^{-1}\begin{bmatrix} 1+\tfrac{1}{4}\left(a^2-b^2-c^2\right) & -c+\tfrac{1}{2}ab & b+\tfrac{1}{2}ac \\ c+\tfrac{1}{2}ba & 1+\tfrac{1}{4}\left(-a^2+b^2-c^2\right) & -a+\tfrac{1}{2}bc \\ -b+\tfrac{1}{2}ca & a+\tfrac{1}{2}cb & 1+\tfrac{1}{4}\left(-a^2-b^2+c^2\right) \end{bmatrix}$$

where $\Delta = 1 + \frac{1}{4}(a^2 + b^2 + c^2)$.

Another algebraic formulation of the rotation matrix is the *Pope–Hinsken matrix* which has practical application in resection (section 2.2.4):

$$\mathbf{R} = \begin{bmatrix} d^2+a^2-b^2-c^2 & 2(ab+cd) & 2(ac-bd) \\ 2(ab-cd) & d^2-a^2+b^2-c^2 & 2(bc+ad) \\ 2(ac+bd) & 2(bc-ad) & d^2-a^2-b^2+c^2 \end{bmatrix}$$

where $a^2 + b^2 + c^2 + d^2 = 1$. (Only three of the parameters are independent.)
The rotation matrix can also be expressed in terms of the cosines of the angles between

the primary axes (XYZ) and the secondary axes (xyz):

$$\mathbf{R} = \begin{bmatrix} \cos(Xx) & \cos(Xy) & \cos(Xz) \\ \cos(Yx) & \cos(Yy) & \cos(Yz) \\ \cos(Zx) & \cos(Zy) & \cos(Zz) \end{bmatrix}.$$

2.1.2 Special cases

A common occurrence in close range photogrammetry is the use of rotations that are integer multiples of $\pi/2$ radians, or close to those values. In such cases, a preliminary transformation matrix might be

$$\mathbf{R} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

which has the effect of interchanging the Y and y axes and the Z and z axes and reversing the directions of the X and x axes.

If the rotation angles are small, so that to first-order accuracy $\sin \Delta\omega = \Delta\omega$ and $\cos \Delta\omega = 1$ etc., we have the skew-symmetric matrix

$$\mathbf{R} = \begin{bmatrix} 1 & \Delta\kappa & -\Delta\varphi \\ -\Delta\kappa & 1 & \Delta\omega \\ \Delta\varphi & -\Delta\omega & 1 \end{bmatrix}.$$

Further discussion of rotation matrices can be found in Shih (1990).

2.1.3 Three dimensional conformal transformation

If the origin of primary (XYZ) axes is translated by (X_o, Y_o, Z_o) the scale along each axis is multiplied by λ and the axes given sequential rotations ω, φ and κ to give secondary (xyz) axes (Figure 2.3) the relationship between the primary co-ordinates (X, Y, Z) of a point A and the secondary co-ordinates (x, y, z) can be expressed by the vector equation

$$\mathbf{X} = \mathbf{X}_o + \lambda^{-1}\,\mathbf{R}^t\,\mathbf{x}. \tag{2.7a}$$

The transformation of the secondary co-ordinates to primary co-ordinates in full is

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix} + \lambda^{-1} \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{2.7b}$$

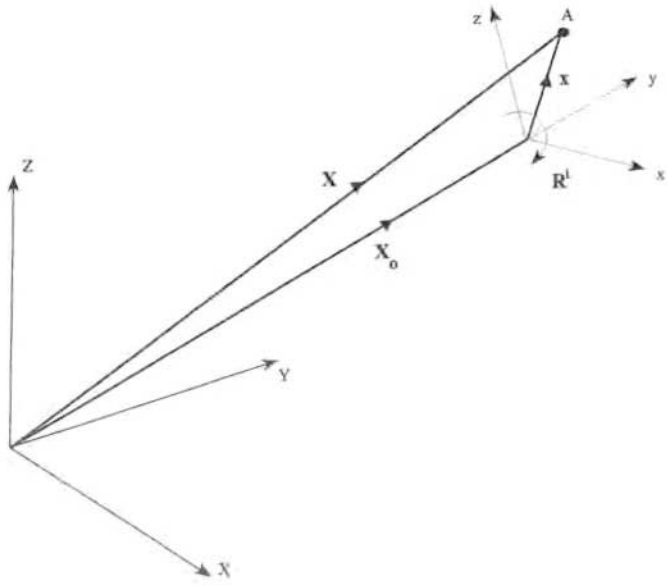

Figure 2.3 *Conformal transformations.*

where the elements of $\mathbf{R}^t$ are defined by (2.4) and (2.5). The inverse transformation (primary to secondary) is $\mathbf{x} = \lambda\mathbf{R}(\mathbf{X} - \mathbf{X}_o)$. Seven parameters ($X_o$, Y_o, Z_o, ω, φ, κ and λ) define the transformation. It is described as *conformal* because the shape of an object defined by co-ordinates in one system is unchanged by the transformation. It is sometimes called a *similarity transformation*.

2.1.4 Two dimensional transformations

In order to take account of physical effects in the image plane, it is often necessary to make co-ordinate transformations that are not conformal. The plane conformal transformation of co-ordinates (x, y) in a secondary (xy) system to co-ordinates (X, Y) in the primary system is

$$X = X_o + \lambda^{-1}(x\cos\theta - y\sin\theta) \tag{2.8a}$$

$$Y = Y_o + \lambda^{-1}(x\sin\theta + y\cos\theta) \tag{2.8b}$$

where the primary origin is translated to (X_o, Y_o) the axes are scaled by λ and rotated anti-clockwise through an angle θ. In matrix notation the transformation can be written as

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X_o \\ Y_o \end{bmatrix} + \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{2.9}$$

where the (2×2) matrix is orthogonal with $a = \lambda^{-1}\cos\theta$ and $b = \lambda^{-1}\sin\theta$. It follows that $\lambda = (a^2 + b^2)^{-\frac{1}{2}}$ and that $\tan\theta = (b/a)$. The transformation has four independent parameters.

If the scaling of the X and Y axes is not the same, but is λ_X along the X-axis and λ_Y along the Y-axis, the transformation of primary co-ordinates to secondary is

$$x = (X - X_o)\lambda_X \cos\theta + (Y - Y_o)\lambda_Y \sin\theta \tag{2.10a}$$

$$y = -(X - X_o)\lambda_X \sin\theta + (Y - Y_o)\lambda_Y \cos\theta \tag{2.10b}$$

and in matrix notation the transformation of secondary co-ordinates to primary is

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X_o \\ Y_o \end{bmatrix} + \begin{bmatrix} \lambda_X^{-1}\cos\theta & -\lambda_X^{-1}\sin\theta \\ \lambda_Y^{-1}\sin\theta & \lambda_Y^{-1}\cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

The transformation has five independent parameters. The (2×2) matrix for scale and rotation is not orthogonal if λ_X and λ_Y are different. A two dimensional transformation more commonly used in photogrammetry is the *affine* transformation with six independent parameters:

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X_o \\ Y_o \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{2.11}$$

The affine transformation introduces a translation of origin, different scale changes along each axis and different rotations of each axis.

More general two dimensional co-ordinate transformations can be expressed as polynomials of the form

$$X = a_o + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2 + a_6xy^2 + ... \tag{2.12a}$$

$$Y = b_o + b_1x + b_2y + b_3xy + b_4x^2 + b_5y^2 + b_6xy^2 + ... \tag{2.12b}$$

where the degree of the polynomial and the non-zero coefficients will be chosen to suit particular circumstances. A commonly used polynomial transformation is known as the *bilinear transformation*:

$$X = a_o + a_1x + a_2y + a_3xy \tag{2.13a}$$

$$Y = b_o + b_1x + b_2y + b_3xy. \tag{2.13b}$$

2.2 Single camera geometry

The starting point for building a functional model for close range photogrammetry is the *central perspective projection* illustrated in Figure 2.4. A point A in the three dimen-

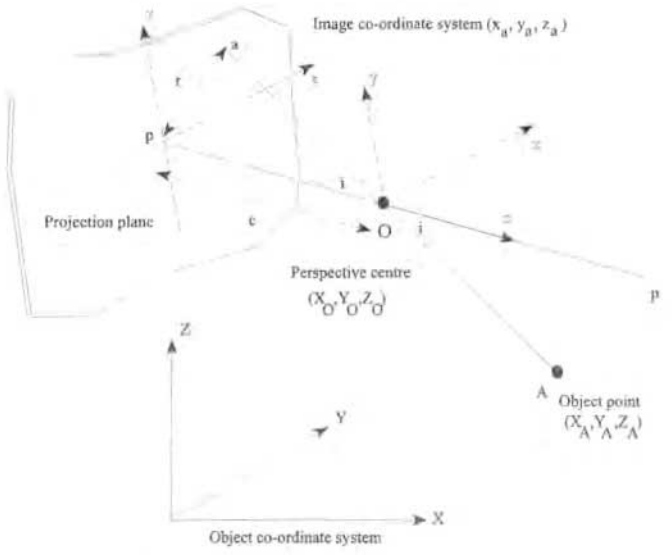

Figure 2.4 *The central perspective projection.*

sional object space is projected onto the *projection plane* by the straight line AOa from A passing through the *perspective centre* O. The *perspective axis* Pop is orthogonal to the projection plane which it intersects at p, the *principal point*. The distance Op from the perspective centre to the plane of projection is the *principal distance*, usually denoted by c. Points A and a are *homologous points*.

Two three dimensional cartesian co-ordinate systems are now introduced in order to derive functional relationships between the position of object point A and the position of a, the projection of A. The primary co-ordinate system (XYZ) is located arbitrarily in the object space. In this system the co-ordinates of the perspective centre of the secondary system are (X_o, Y_o, Z_o) and the co-ordinates of point A are (X_A, Y_A, Z_A). The secondary cartesian system (xyz) has its origin at O, the perspective centre; its z-axis coincides with the principal axis and is directed away from the plane of projection; its x- and y-axes are parallel to the the plane of projection and are directed so as to complete a right-hand system; the scale factor λ is unity. The co-ordinates of point a in the secondary system are (x_a, y_a, $-c$). Figure 2.5 combines elements of Figures 2.3 and 2.4 and shows the collinearity of A, O and a. If the vectors are written relative to the primary (object space) co-ordinate system: $\mathbf{X}_A = \mathbf{X}_O + \mathbf{S}$ where $\mathbf{S}$ is the position vector of A relative to O. It is collinear with $\mathbf{x}_a$, but of opposite sense. Therefore

$$\mathbf{X}_A = \mathbf{X}_O - \mu \mathbf{R}^t \mathbf{x}_a \tag{2.14a}$$

which in matrix notation is

$$\begin{bmatrix} X_A \\ Y_A \\ Z_A \end{bmatrix} = \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix} - \mu \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix} \begin{bmatrix} x_a \\ y_a \\ -c \end{bmatrix} \tag{2.14b}$$

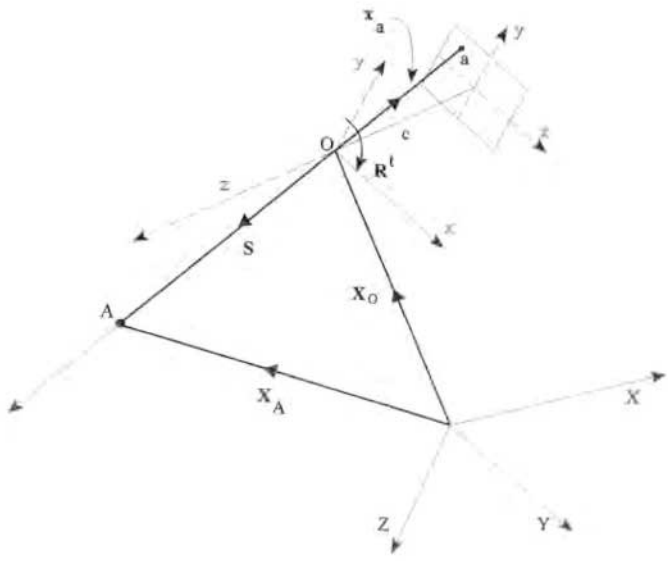

Figure 2.5 *The collinearity condition.*

where μ is a scalar quantity greater than zero and the elements r_{ij} of the rotation matrix **R** are given by (2.4) and (2.5). Equations (2.14a) and (2.14b) illustrate the fact that if the position of the perspective centre, the direction of the perspective axis and the principal distance are known, there is no unique solution for the co-ordinates (X_A, Y_A, Z_A) of A, given the position of its projection a; the direction of A relative to the perspective centre is known, but the parameter μ is indeterminate without more information (the distance of A from O, for example).

The reverse transformation (primary co-ordinates to secondary) is

$$\mathbf{x}_a = \mu^{-1}\mathbf{R}(\mathbf{X}_O - \mathbf{X}_A) \tag{2.15a}$$

which in matrix notation is

$$\begin{bmatrix} x_a \\ y_a \\ -c \end{bmatrix} = \mu^{-1}\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}\begin{bmatrix} X_A - X_O \\ Y_A - Y_O \\ Z_A - Z_O \end{bmatrix}. \tag{2.15b}$$

The third equation of (2.15b) can be written explicitly in μ^{-1} and the expression substituted in the first two equations of (2.15b) to give

$$x_a = \frac{-c\left[r_{11}(X_O - X_A) + r_{12}(Y_O - Y_A) + r_{13}(Z_O - Z_A)\right]}{\left[r_{31}(X_O - X_A) + r_{32}(Y_O - Y_A) + r_{33}(Z_O - Z_A)\right]} \tag{2.16a}$$

$$y_a = \frac{-c\left[r_{21}(X_O - X_A) + r_{22}(Y_O - Y_A) + r_{23}(Z_O - Z_A)\right]}{\left[r_{31}(X_O - X_A) + r_{32}(Y_O - Y_A) + r_{33}(Z_O - Z_A)\right]}. \qquad (2.16b)$$

These equations are referred to by photogrammetrists as *collinearity equations*; they are derived from the collinearity of a point such as A, the perspective centre O and the perspective projection of A onto the plane of projection at a. If the plane of projection is taken to lie on the same side of the perspective centre as A then the negative signs of c in (2.16a) and (2.16b) become positive.

2.2.1 Camera calibration

The central perspective projection is an abstract entity. When it is used in photogrammetry as the mathematical model for image formation in a camera, significant discrepancies leading to errors will occur. Photogrammetrists seek to identify why and by how much the geometry of image formation in a real camera differs from the geometry of a central perspective projection. Such a procedure is known as *camera calibration*. The most obvious discrepancy between the abstract and the real is the lens. It is represented in Figures 2.4 and 2.5 as a point whereas Figure 2.6 shows a longitudinal section through an asymmetric lens that has front and rear elements with an aperture stop S between them for controlling the amount of light passing to the image plane and for maintaining depth of field. If the lens is at infinity focus, a bundle of parallel rays from a distant target A will pass through the optical system as shown and will be brought to a focus in the image plane at a. The image plane is the realisation, in a camera, of the plane of projection in a perspective projection. Rays from a second target B in the same direction as A, but closer to the camera, will come to a focus at b, behind the image plane. The image of B at the image plane is the *blur circle* which can be reduced to the *circle of least confusion* by decreasing the diameter of the aperture stop.

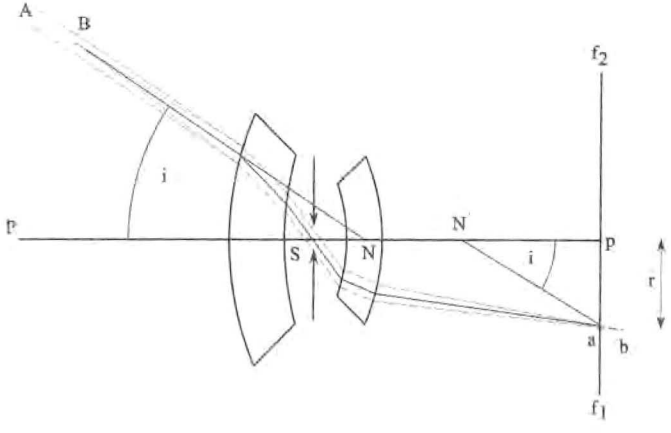

Figure 2.6 *Image formation by a lens (after Scott, 1977).*

If a camera is to be treated as geometrically the same as a perspective projection, two perspective centres can be defined: N and N' respectively in Figure 2.6. For a lens focused at infinity, these points are two cardinal points of unit angular magnification (nodal points). The front nodal point N is defined by the intersection of the chief or principal ray, incident at an angle i, with the optical axis of the camera. The rear nodal point is defined by the intersection with the optical axis of a line from the image a making an angle i with the the optical axis. At infinity focus, the position of the inner node depends upon the angle of incidence. Such variation in the principal distance is usually modelled as radial lens distortion (*see below* at 2.2.2) and the principal distance for infinity focus is defined for paraxial rays only. For a change δc in c, the corresponding change in r is $\delta r = (r\delta c)/c$. Radial lens distortion should be determined by calibration.

Unfortunately this perspective centre definition holds only for objects at infinity. Scott (1977) shows that, at reduced object distances, the rear perspective centre coincides not with the rear node, but with the exit pupil, defined as the image of the aperture stop seen from the image space. This means that when the camera lens is moved away from the film plane by a physical extension δc to facilitate the focusing of a close object, the principal distance for infinity focus will not only be increased by δc, but also by an extra increment $\delta c'$, which varies with radial distance, due to variation in location of the exit pupil. The variation of principal distance with focus setting must be determined through photogrammetric calibration of the camera, since the exit pupil position is very difficult to locate in practice. The modelling of Δc and of $\Delta c'$ can again be done through assigning corresponding values to δr. An additional complication occurs when object distances are within a few tens of focal lengths from the camera. The position of the outer perspective centre varies with object distance (Magill, 1955) and the corresponding variation in the principal distance δc can again be modelled by means of changes δr.

A camera that has been designed and constructed to be used for photogrammetry is a *metric camera*. Traditionally such designs have included specific features to ensure close conformance to the perspective projection model. The lens is designed to ensure that as far as possible for any particular focal setting the position of the rear node is independent of the angle of incidence, in other words that the principal distance is constant for all incident rays for a given focal setting. Moreover, each time a particular focal setting is repeated, the corresponding principal distance should be ensured by well-engineered optical and mechanical components. Lens elements are designed to be coaxial and to remain so during focusing so that the optical axis Pp is orthogonal to the image plane and intersects it at the point defined by the intersection of the lines between pairs of collimating points, or *fiducial marks* such as f_1 and f_2 in Figure 2.6. The fiducial marks are fixed to the camera body and lie in the image plane. They define the *fiducial axes* which are designed to be orthogonal and which are the datum for definition of positions of points in the image.

Even though a metric camera is designed to produce imaging geometry that is a realization of the perspective centre projection, imperfections are inevitable so calibration is necessary. Manufacturers provide a calibration certificate for each metric camera produced. The minimum calibration data should be the calibrated principal distances c

corresponding to specific focal settings, and the co-ordinates (x_o, y_o) of the the principal point with respect to the fiducial axes (sometimes called the *principal point offset*). These three elements are the *interior orientation* elements of the camera. Usually the principal distance value is quoted to the nearest 10 μm. Metric camera construction methods usually ensure that the principal point is offset from the intersection of the fiducial axes by no more than a few μm in either direction. Once the co-ordinates (x_o, y_o) are known, the fiducial co-ordinates (x', y') of an image can be reduced to the principal point to give photoco-ordinates $x = (x' - x_o)$ and $y = (y' - y_o)$. Co-ordinates of the fiducial marks are also usually given on the manufacturer's calibration certificate. Calibrated values vary with camera usage, so regular re-calibration should be routine.

Whilst there is no substitute for precision engineering in good photogrammetric camera design, the technological drive towards digital imaging combined with advances in analytical and digital photogrammetric techniques mean that traditional photogrammetric camera requirements are being increasingly met by other means. For example fiducial marks are not necessary because the regular array of light sensitive elements within the CCD image sensor can be used as a photoco-ordinate system. Lenses optimized for pictorial photography, often with short focal lengths, are being routinely used. These often have large lens distortions and principal point offsets which must be measured by a combination of laboratory techniques and self-calibration bundle adjustments. Parameters from these estimations can then be included in subsequent photogrammetric data processing.

2.2.2 Radial and tangential lens distortion

Variations in angular magnification with angle of incidence are usually interpreted as *radial lens distortion*. Design and construction of metric cameras ensure that such distortion is minimized, but calibration is usually necessary. Radial lens distortion is usually expressed as a polynomial function of the radial distance from the *point of symmetry*, which usually coincides with the principal point:

$$\delta r = K_1 r^3 + K_2 r^5 + K_3 r^7$$

where δr is the radial displacement of an image point, $r^2 = (x - x_o)^2 + (y - y_o)^2$, (x, y) are the fiducial co-ordinates of the image point, (x_o, y_o) are the fiducial co-ordinates of the point of symmetry, commonly the principal point, and K_1, K_2 and K_3 are coefficients whose values depend upon the camera focal setting. The distortion δr is usually resolved into two components:

$$\delta r_x = \delta r(x - x_o)/r \qquad (2.17a)$$

and

$$\delta r_y = \delta r(y - y_o)/r \qquad (2.17b)$$

For a large format metric camera (up to 150mm say) the maximum magnitude of radial lens distortion is about 10–20 μm. Typical radial lens distortion curves for some inexpensive CCD camera lenses are illustrated in Figure 2.7.

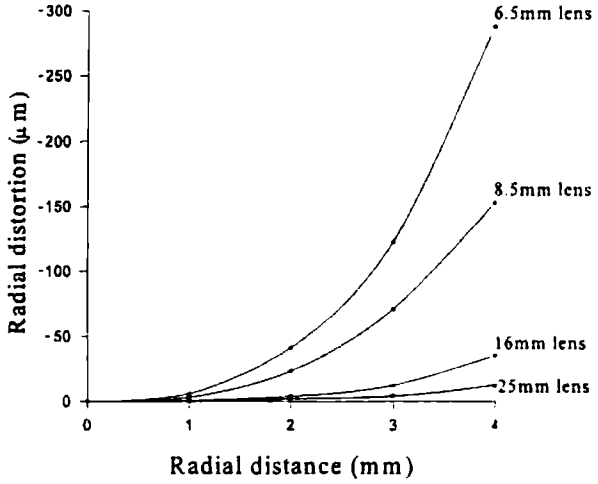

Figure 2.7 *Typical radial lens distortion curves for 25, 16, 8·5 and 6·5 mm C mount lenses fitted to a Pulnix TM6CN camera.*

Tangential lens distortion is the displacement of a point in the image caused by misalignment of the components of the lens. The displacement is described by two polynomials, one for the displacement in the direction of the x-fiducial axis and the other for displacement in the y-direction:

$$\delta x = P_1[r^2 + 2(x - x_o)^2] + 2P_2(x - x_o)(y - y_o) \qquad (2.18a)$$

$$\delta y = P_2[r^2 + 2(y - y_o)^2] + 2P_1(x - x_o)(y - y_o) \qquad (2.18b)$$

where P_1 and P_2 are coefficients whose values depend on the camera focal setting. The other terms are the same as those in (2.17a) and (2.17b).

2.2.3 Image refinement

In a traditional metric camera, the realization of the plane of projection by an image plane is made either by using glass plates carrying a photographic emulsion, or by a film flattening device such as a glass plate or air pump. Neither of these is completely satisfactory, so some analytical methods for correcting the geometric effects on the image of its lack of flatness and non-orthogonality to the principal axis must be developed. Furthermore, differential scale change may be introduced in an image during chemical processing of the emulsion. Correction of photoco-ordinates for these effects is known as *image refinement*. Before this can be carried out, the two dimensional image co-ordinate system must be defined. In metric cameras that have either a film flattening device or make use of glass plates, this is usually done by fiducial marks located in the image plane and integral with the camera body which define a pair of orthogonal axes; images of these fiducial marks on the photographs are produced either by incident light through the

objective or by independent illumination. In some metric cameras the image co-ordinate system is defined by a réseau of crosses on a glass plate fixed in the camera body at the image plane. The glass plate can also help to keep the film flat during exposure.

Some non-metric photographic cameras can be adapted by fixing a glass réseau plate in the image plane to help film flattening and to provide an image co-ordinate system. They may also be modified by mechanically fixing the lens and its focal setting, or by defining click-stops in the mechanical lens focusing system. Conventional photographic cameras modified in this way are sometimes called *semimetric* cameras. Image co-ordinates can also be defined by the edges of the format frame, the procedure depending on their regularity and definition in the photographic image. Elements necessary to define two dimensional co-ordinates are a straight line and a point; the line defines one axis (the other is defined as orthogonal to it) and the point defines the origin. In any particular camera, the line could be defined as a best fit to one of the (longer) edges of the format frame and the point could be the intersection of this line with the line which is a best fit to one of the shorter edges (which will probably not be orthogonal to the line through the longer edge). In an image from a CCD camera the rectangular array of rows and columns of pixels is usually used as an implicit rectangular co-ordinate system, the integer numbers of the row and column of the pixel (or integer and fraction for sub-pixel location) relative to a specific corner of the array are the image co-ordinates. When the true origin is in or near a corner of the format, a false origin near the centre can be defined for convenience as in Figure 2.8.

Images from photographic cameras are usually placed in a comparator where two dimensional *comparator co-ordinates* of images of fiducial or réseau marks, or edges of format frames, and targets are measured. Measurement may be carried out by an operator, sometimes making use of automated processes, or by fully automatic methods. Comparator co-ordinates are transformed to image co-ordinates, then, by applying the

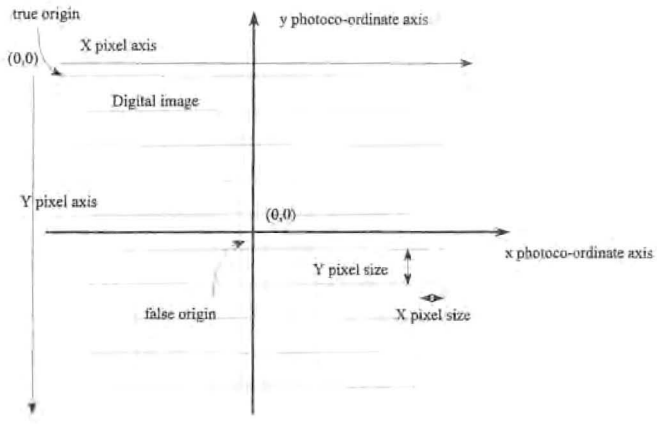

Figure 2.8 *Co-ordinate systems for photoco-ordinates of a digital image.*

principal point offset, to photoco-ordinates which may then be refined for various reasons before applying corrections for lens distortions if relevant. For a metric camera with the usual four fiducial marks, measured comparator co-ordinates (X, Y) of the fiducial marks can be used in conjunction with their calibrated image co-ordinates (x, y) to evaluate up to eight transformation parameters. Usually one of the following is chosen: conformal transformation (2.9) either with unit scale factor and three parameters to be evaluated or with four parameters to be evaluated; and affine transformation (2.11) with six parameters. The latter allows for non-orthogonality of comparator axes. If a calibrated réseau is used in order to refine the photo co-ordinates for lack of image flatness, bilinear transformations (2.13a) and (2.13b) based on the four réseau marks surrounding each target image can be made. In rare cases the use of a more general polynomial for refining photoco-ordinates might be used, but the time and trouble involved in identifying relevant terms might not be justified by the improvement in accuracy.

It is usual to have redundant measurements for the evaluation of the particular transformation parameters, so the use of least squares estimation processes (section 2.5) is appropriate. The transformation formulae are used as the functional model (2.33) in which the transformation parameters are $\mathbf{x}$, the unknowns to be estimated, and the comparator co-ordinates are the measurements $\mathbf{b}$, usually with the identity matrix as the a priori weight matrix. The calibrated values of the co-ordinates of the fiducial and réseau marks are usually regarded as constants; the inconvenience in treating them rigorously as correlated random variables is normally not justified. Comparator co-ordinates of points along the image of an edge of the format frame in a non-metric camera can be used to derive the equation of a best fitting straight line and subsequently to transform comparator co-ordinates to image co-ordinates. It is likely that the image of a format edge on one photograph from a non-metric camera will not be exactly the same as its image on another photograph. The image co-ordinate datum will therefore be slightly different for each photograph taken with the camera so inaccuracies will occur. They will be increased by lack of information about film unflatness provided by a réseau.

For digital images, automated methods for locating the centre of a target are based on analyses of the two dimensional intensity distribution. The position is usually defined in terms of pixel row and column numbers, to sub-pixel resolution. The most important image refinement is to account for scale differences along the two pixel directions arising from the dimensions of the CCD elements (Figure 2.8) and physical processes involved in analogue to digital conversion, registration and storage. In the absence of an independent calibration, the scale discrepancy can be allowed for as an additional parameter in the bundle adjustment (section 2.4).

Having defined an image co-ordinate datum (by fiducial marks, or réseau, or format edges, or pixel numbers) co-ordinates in that datum (x', y') must be transformed by the principal point offset (x_o, y_o) to $x = (x' - x_o)$ and $y = (y' - y_o)$ to give photoco-ordinates.

2.2.4 Resection

Measurements of photoco-ordinates x_a and y_a (refined and corrected for lens distortion if relevant) of the image a of target A (Figure 2.5) give rise to two collinearity equations

(2.16a) and (2.16b). If the three elements of interior orientation (c, x_o and y_o) are given by camera calibration and co-ordinates (X_A, Y_A, Z_A) of A in the object space co-ordinate system are known (from measurements in the object space for example) the two equations will have six unknowns: rotations ω, φ and κ and co-ordinates (X_o, Y_o, Z_o) of the perspective centre. These six parameters are elements of *exterior orientation*. Their evaluation is known as *resection*. At least three non-collinear targets such as A (*control points*) are necessary for resection of a camera. The method used to evaluate the elements of exterior orientation depends upon the purpose of the photogrammetry and whether speed or a statistically rigorous estimation is more important.

A method for the direct evaluation of the six elements of exterior orientation from measured photo co-ordinates of images of three non-collinear control points that does not require any approximate values is described by Zeng and Wang (1992). The procedure gives the co-ordinates of the perspective centre directly. An algebraic form of rotation matrix is used (the Pope–Hinsken matrix, section 2.2.1). If needed, values of the three rotations ω, φ and κ can be found from the nine elements of the rotation matrix by the procedure described in section 2.2.

If a more statistically rigorous resection is required, the collinearity equations (2.16a) and (2.16b) can be linearized and an iterative least squares estimation process used (section 2.5) to evaluate the six exterior orientation elements. In this case, more than three control points can (and should) be used. Additionally, if enough control points are available and are well distributed throughout the object space, it may be possible to make accurate estimates of interior orientation parameters as well. Different precisions of control co-ordinates can be included and consequent statistical properties of the estimated orientation elements can be used in further photogrammetric processing, such as intersection. The disadvantage of the rigorous resection is the need to derive approximate values for the orientation elements that are sufficiently close to the final values for the iterative process to converge. Usually, appropriate values for the co-ordinates (X_o, Y_o, Z_o) can be readily obtained, but that is not so for values of the rotation angles; a direct solution using three non-collinear control points as described above will give good initial values for the statistically rigorous procedure.

Resection is only an intermediate stage in a photogrammetric procedure; it is often followed by intersection (*see below*) or by a rigorous multistation bundle adjustment (section 2.5) which uses the values of the exterior orientation elements as approximate initial values.

2.3 The geometry of two cameras

2.3.1 *Intersection*

If the elements of exterior orientation of two cameras with perspective centres at O_1 and O_2 (Figure 2.9) are known (from two independent resections for example) the object space co-ordinates (X_A, Y_A, Z_A) of a target A can be evaluated from measurements of the photo co-ordinates (x_1, y_1) and (x_2, y_2) of its homologues a_1 and a_2. Collinearity equations (2.16a) and (2.16b) are the basis of the method, but because there are three unknowns (X_A, Y_A, and Z_A) and four equations, a least squares estimation is usually made.

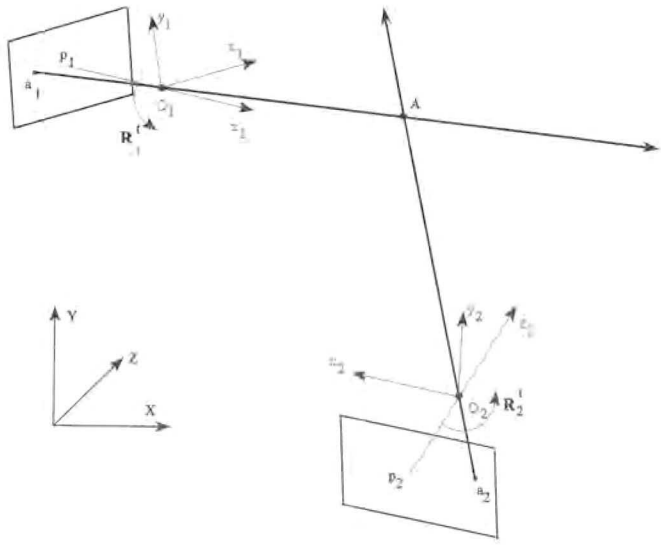

Figure 2.9 *Intersection.*

The process is known as *intersection*. As with resection, the intersection procedure used will depend upon whether speed or statistical rigour is more important. A linear least squares estimation of the co-ordinates of a target such as A can be derived by regarding all terms (except X_A, Y_A and Z_A) in the collinearity equations (2.16a) and (2.16b) as known constants. Then (2.47) with $\mathbf{W} = \mathbf{I}_4$ and $\mathbf{b}$ as a constant vector can be used to evaluate X_A, Y_A and Z_A. Another intersection procedure that is direct is described in the next sub-section and by (2.23).

Independent resections of cameras followed by intersections of object space targets using pairs of cameras and homologous points are common procedures in close range photogrammetry. Such procedures do not, however, make full use of the opportunities for accurate, precise and reliable estimates of object space co-ordinates of targets that close range photogrammetry offers. A multistation convergent bundle adjustment (sections 2.4 and 2.5) gives results of higher quality (section 2.6) than resection and intersection, but at a cost. In a bundle adjustment all measured photoco-ordinates are processed simultaneously, using (iterative) linearized least squares methods, to evaluate exterior (and sometimes interior) orientation elements, camera calibration data and object space co-ordinates. The simultaneous solution involves large data sets and the computation of inverses of large matrices. Resection and intersection on the other hand are much quicker; in the former the inverses of a series of (6 × 6) matrices are needed, in intersection the dimensions are only (3 × 3).

However, only a few degrees of freedom are available (in the case of an intersection there is only one degree of freedom) so the results are of low reliability. In a multistation bundle adjustment, a few thousand measurements and several hundred unknowns are not uncommon, so the number of degrees of freedom is often high.

A further advantage of the multistation bundle adjustment now that digital cameras are increasingly used in close range photogrammetry, is that it permits the evaluation of elements of camera calibration and interior orientation to be made far more reliably than is possible using resection and intersection. In resection, the large number of control points needed to provide reliable and accurate estimates of camera calibration and interior orientation elements as well as the elements of exterior orientation might be impossible to obtain quickly and economically.

If statistical rigour in intersection is important, estimates of exterior orientation elements and their covariance matrix from resections (or from a multistation bundle adjustment) can be used with linearized collinearity equations for a series of iterative least squares estimations of target co-ordinates based on (2.54). This procedure relies on the exterior orientation data being still valid for the intersection in progress. In industrial high precision photogrammetry, a camera configuration around a calibration field could be used in a multistation bundle adjustment which would provide exterior orientation and other data. The calibration field could then be removed and replaced by an object to be measured. It might be appropriate to assume that the camera orientation and calibration data remain valid so that the co-ordinates of the new targets can be derived by a series of iterative intersections using the statistical data deemed necessary. Examination of the quality of data (section 2.6) should indicate whether the camera orientation and calibration data being used are still valid or whether a new estimation using the calibration field is needed. Direct methods of intersection could be used to obtain approximate values of the target co-ordinates.

2.3.2 Coplanarity equation

Evaluation of the exterior orientation elements of one camera relative to the photo co-ordinate system of another is known as *relative orientation*. In Figure 2.10, suppose photoco-ordinate axes $(x_1 y_1 z_1)$ of the image having its perspective centre at O_1 are primary co-ordinates and it is required to evaluate the exterior orientation elements of a second camera with its perspective centre at O_2. Target A is imaged at a_1 and at a_2. Vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{b}$ are coplanar and lie in the *epipolar plane* of target A and the two perspective centres. Relative to the primary axes these vectors are:

$$\mathbf{a}_1 = -\lambda[x_1 \;\; y_1 \;\; -c_1]^t ; \qquad \mathbf{a}_2 = -\mu \mathbf{R}^t [x_2 \;\; y_2 \;\; -c_2]^t ; \qquad \mathbf{b} = [b_x \;\; b_y \;\; b_z]^t$$

where $\mathbf{b}$ is the *camera base* and $\mathbf{R}^t$ is given by (2.6b). The triple scalar product of three coplanar vectors is zero, so $\mathbf{b}.\mathbf{a}_1 \times \mathbf{a}_2 = 0$, or

$$\det \begin{bmatrix} b_x & x_1 & r_{11}x_2 + r_{21}y_2 - r_{31}c_2 \\ b_y & y_1 & r_{12}x_2 + r_{22}y_2 - r_{32}c_2 \\ b_z & -c_1 & r_{13}x_2 + r_{23}y_2 - r_{33}c_2 \end{bmatrix} = 0 . \qquad (2.19)$$

Assuming that at least one of the components (say b_y) of the base vector $\mathbf{b}$ is non-zero, and putting $\mathbf{R}^t \mathbf{a}_2 = \mathbf{a}_2' = [x_2' \;\; y_2' \;\; z_2']^t$:

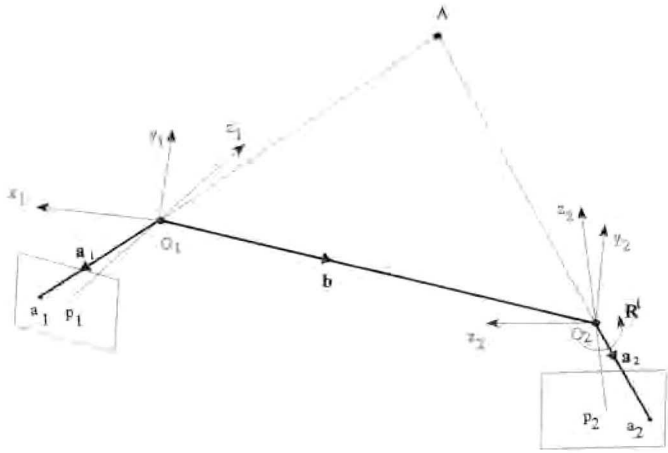

Figure 2.10 *Coplanarity.*

$$\det\begin{bmatrix} 1 & x_1 & x_2' \\ \frac{b_y}{b_x} & y_1 & y_2' \\ \frac{b_z}{b_x} & -c_1 & z_2' \end{bmatrix} = 0. \qquad (2.20)$$

This equation is the *coplanarity equation* for target A. There are five elements of relative orientation: (b_y/b_x), (b_z/b_x), ω, φ and κ. Measured photoco-ordinates of the images of a minimum of three targets, well distributed throughout the object space, can be used to evaluate the five elements of relative orientation, but it is usual to measure more than the minimum and evaluate the five elements by linear least squares estimation. Equation (2.20) includes photoco-ordinates (x_1, y_1) and (x_2, y_2) for each target image as measured elements, principal distances c_1 and c_2 as elements with prior estimates and (b_y/b_x), (b_z/b_x), ω, φ and κ as elements to be estimated. It is therefore an example of the generalized least squares problem (section 2.5) solved rigorously using (2.42) and (2.43). However, it is customary to assume that measured values of (x_1, y_1) and (x_2, y_2) and the prior estimates of c_1 and c_2 are constants and to treat (2.20) as a linear equation in five unknowns, with the least squares estimate obtained from (2.47) with $\mathbf{W} = \mathbf{I}$ and $\mathbf{b}$ as a constant vector.

After relative orientation, it is possible to use measured photoco-ordinates of the two homologues of a target in order to evaluate, by intersection, its co-ordinates relative to the $(x_1 y_1 z_1)$ axes. These co-ordinates are usually referred to as *model co-ordinates*, the name deriving from the stereoscopic model commonly used in traditional aerial mapping procedures. An algorithm for the intersection computation that is different from the one given in the previous sub-section is based on Figure 2.11 in which $\mathbf{s}$ and $\mathbf{t}$ are

unit vectors in directions opposite to $\mathbf{a}_1$ and $\mathbf{a}_2$ respectively and, in relation to the $(x_1y_1z_1)$ axes, are defined by:

$$\mathbf{s} = -\mathbf{a}_1 / |\mathbf{a}_1| \text{ and } \mathbf{t} = -\mathbf{a}_2' / |\mathbf{a}_2'|.$$

Figure 2.11 illustrates the fact that there is no unique intersection of vectors $\lambda\mathbf{s}$ and $\mu\mathbf{t}$, where λ and μ are scalars greater than zero, because in practice $\mathbf{s}$ and $\mathbf{t}$ are derived from inexact measured photoco-ordinates of a_1 and a_2, calibrated principal distances c_1 and c_2 and estimates of the values of exterior relative orientation elements of the camera at O_2. Vectors $\lambda\mathbf{s}$ and $\mu\mathbf{t}$ are skew; $\mathbf{p}$ is any vector from $\lambda\mathbf{s}$ to $\mu\mathbf{t}$ and is known as the *parallax* vector. It is given by

$$\mathbf{p} = -\lambda\mathbf{s} + \mathbf{b} + \mu\mathbf{t} \tag{2.21}$$

In order to define a unique position for target A on $\mathbf{p}$, it is necessary first to specify a property of the parallax vector that will be satisfied by specific values of the scalars λ and μ. For example, it might be appropriate (as it often is in aerial mapping) to require the components p_x and p_z to be zero. The non-zero component of $\mathbf{p}$ is then often referred to as the *residual parallax*, or *y-parallax*. In close range photogrammetry it is more usual to require the parallax vector to have minimum length and then to define the position of target A as its mid-point. If $\mathbf{p}$ is to have minimum length,

$$\frac{\partial|\mathbf{p}|}{\partial\lambda} = \frac{\partial|\mathbf{p}|}{\partial\mu} = 0 \tag{2.22}$$

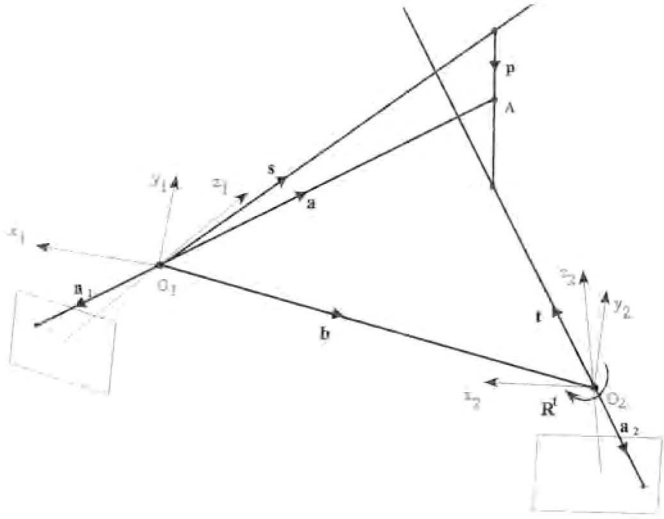

Figure 2.11 *Parallax vector and model co-ordinates.*

which is satisfied when the values of λ and μ are λ_o and μ_o respectively, given by

$$\lambda_o = \frac{-(\mathbf{s}.\mathbf{t})(\mathbf{t}.\mathbf{b}) + (\mathbf{s}.\mathbf{b})}{1-(\mathbf{s}.\mathbf{t})^2}$$

and

$$\mu_o = \frac{(\mathbf{s}.\mathbf{t})(\mathbf{s}.\mathbf{b}) - (\mathbf{t}.\mathbf{b})}{1-(\mathbf{s}.\mathbf{t})^2}$$

recalling that $\mathbf{b} = [1 \;\; b_y/b_x \;\; b_z/b_x]^t$. Note that λ_o and μ_o are infinite if the scalar product $(\mathbf{s}.\mathbf{t})$ is unity, i.e. if $\mathbf{s}$ and $\mathbf{t}$ are parallel. The position vector $\mathbf{a}$ of target A, relative to the co-ordinate axes $(x_1 y_1 z_1)$ is then given by

$$\begin{aligned} \mathbf{a} &= \lambda_o \mathbf{s} + \tfrac{1}{2}\mathbf{p} \\ &= \tfrac{1}{2}(\lambda_o \mathbf{s} + \mathbf{b} + \mu_o \mathbf{t}) \end{aligned} \qquad (2.23)$$

from (2.21). The residual parallax vector

$$\mathbf{p}_o = -\lambda_o \mathbf{s} + \mathbf{b} + \mu_o \mathbf{t}$$

is an indicator of the quality of the positioning of A. This intersection procedure is direct, not iterative. The position vectors of all targets such as A can if necessary be

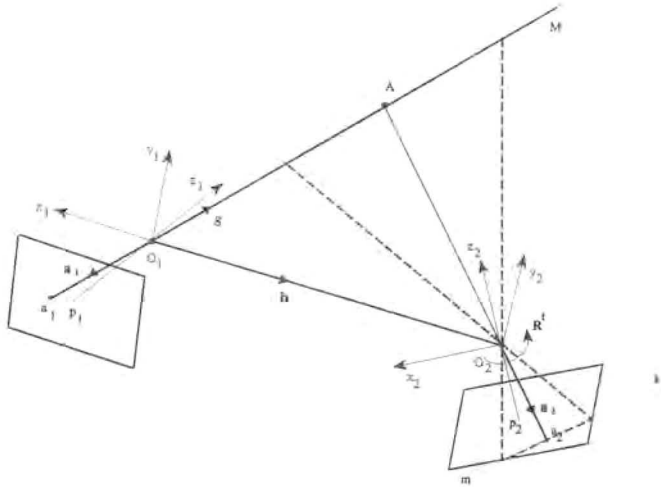

Figure 2.12 *Epipolar plane and an epipolar line.*

transformed to an object space co-ordinate system using a seven parameter three dimensional conformal transformation (section 2.1).

2.3.3 Epipolar geometry

Target A with image a_1 in a camera with perspective centre at O_1 will also give an image a_2 in the camera at O_2 (Figure 2.12). The epipolar plane is shown shaded. It intersects the image plane of the camera at O_2 in the *epipolar line km*, where k is the projection of O_1 through O_2 and M is any point along the vector $\lambda\mathbf{s}$. Assuming that the exterior orientation parameters of the camera at O_2 have been determined relative to the co-ordinate axes $(x_1y_1z_1)$ by the relative orientation procedure described in the preceding section, it is possible to derive the equation of the line km relative to the co-ordinate axes $(x_2y_2z_2)$. Thus, given the location of the homologue of A in the camera image at O_1, the homologue of A in the camera image at O_2 will lie along this epipolar line. Any search procedure to locate a_2 as a probable match of a_1 can then be confined to the epipolar line; in practice, because none of the values used to derive the equation of the line is exact, the search takes place along a narrow band centred on the epipolar line.

The image co-ordinates of k are found by using the collinearity equations (2.16a) and (2.16b). From the relative orientation, $\mathbf{b} = [1 \;\; b_y/b_x \;\; b_z/b_x]^t$ and the rotation matrix $\mathbf{R}$ (2.4) and (2.5) are known, so that

$$x_k = \frac{-c_2\left[r_{11} + r_{12}\dfrac{b_y}{b_x} + r_{13}\dfrac{b_z}{b_x}\right]}{\left[r_{31} + r_{32}\dfrac{b_y}{b_x} + r_{33}\dfrac{b_z}{b_x}\right]} \tag{2.24a}$$

and

$$y_k = \frac{-c_2\left[r_{21} + r_{22}\dfrac{b_y}{b_x} + r_{23}\dfrac{b_z}{b_x}\right]}{\left[r_{31} + r_{32}\dfrac{b_y}{b_x} + r_{33}\dfrac{b_z}{b_x}\right]} \tag{2.24b}$$

can be used to evaluate the image co-ordinates of k. Point M can be defined by assigning a value to the positive scalar λ. Then the position vector of M relative to the co-ordinate axes $(x_1y_1z_1)$ is $\lambda\mathbf{s}$, where $\mathbf{s} = -\mathbf{a}_1/|\mathbf{a}_1|$. The co-ordinates of m are therefore

$$x_m = \frac{-c_2\left[r_{11}(1-\lambda s_x) + r_{12}\left(\dfrac{b_y}{b_x} - \lambda s_y\right) + r_{13}\left(\dfrac{b_z}{b_x} - \lambda s_z\right)\right]}{\left[r_{31}(1-\lambda s_x) + r_{32}\left(\dfrac{b_y}{b_x} - \lambda s_y\right) + r_{33}\left(\dfrac{b_z}{b_x} - \lambda s_z\right)\right]} \tag{2.25a}$$

and

$$y_m = \frac{-c_2\left[r_{21}(1-\lambda s_x)+r_{22}\left(\frac{b_y}{b_x}-\lambda s_y\right)+r_{23}\left(\frac{b_z}{b_x}-\lambda s_z\right)\right]}{\left[r_{31}(1-\lambda s_x)+r_{32}\left(\frac{b_y}{b_x}-\lambda s_y\right)+r_{33}\left(\frac{b_z}{b_x}-\lambda s_z\right)\right]} \qquad (2.25b)$$

These x and y co-ordinates of k and m define the two dimensional position vectors **k** and **m** respectively, relative to the principal point of the image. The vector equation of the line through m and k is

$$\mathbf{r} = \mathbf{k} + \nu(\mathbf{m} - \mathbf{k})$$

where ν is a scalar variable. The equation of the epipolar line in the camera at O_1 can be derived in a similar way by projecting O_2 through O_1.

2.3.4 Direct linear transformation

To take advantage of the relatively low cost and widespread availability of non-metric cameras for close range photogrammetry and to obviate the need for fiducial marks or a réseau for image refinement and to give a reference for the position of the principal point, Abdel-Aziz and Karara (1971) devised functional relationships between image co-ordinates and corresponding target co-ordinates that are essentially linear. The method need not require approximate values of the transformation parameters prior to their evaluation. In effect it gives a direct relationship between comparator co-ordinates and object co-ordinates, allowing for differential scale along the image co-ordinate axes. The following two collinearity equations are the basis of the method:

$$x = \frac{L_1X + L_2Y + L_3Z + L_4}{L_9X + L_{10}Y + L_{11}Z + 1} \qquad (2.26a)$$

and

$$y = \frac{L_5X + L_6Y + L_7Z + L_8}{L_9X + L_{10}Y + L_{11}Z + 1} \qquad (2.26b)$$

where (x, y) are image co-ordinates in an arbitrary system (comparator, or CCD array, for example) of a target image, (X, Y, Z) are the object space co-ordinates of the target, and L_1 to L_{11} are the parameters of the transformation and are assumed to be independent. These equations could be regarded as observation equations to be linearized and estimates of the 11 parameters obtained by iterative least squares estimation (section 2.5), but they can be re-written as:

$$xL_9X + xL_{10}Y + xL_{11}Z + x - L_1X - L_2Y - L_3Z - L_4 = 0 \qquad (2.27a)$$

and

$$yL_9X + yL_{10}Y + yL_{11}Z + y - L_5X - L_6Y - L_7Z - L_8 = 0 \qquad (2.27b)$$

so if the measured values (x, y) of image co-ordinates of each target image and object space co-ordinates (X, Y, Z) of each corresponding target are assumed to be error free, these equations can be treated as if they are linear observation equations in the 11 parameters with known coefficients xX, xY, etc. and a direct least squares estimate of the parameters obtained from (2.47) using $\mathbf{W} = \mathbf{I}$ and $\mathbf{b}$ as a constant vector. Subsequently, image co-ordinates can be transformed to object space co-ordinates of corresponding targets provided each target is imaged in two cameras. This can be done by treating (2.27a) and (2.27b) as linear equations with X, Y and Z as parameters to be estimated and image co-ordinates (x, y) and estimated parameters L_1 to L_{11} as constants. Alternatively the object space co-ordinates of a target can be estimated as a general case of iterative least squares processing based on (2.26a) and (2.26b). This will require approximate values of the parameters to be estimated. These can be obtained from three equations like (2.27a) and (2.27b). The comparator co-ordinates can be given appropriate weights in the estimation process.

DLT is a linear treatment of what is essentially a non-linear problem so it gives approximate results. Procedures to make the process more rigorous are discussed by Marzan and Karara (1975) and Bopp and Kraus (1978). Some approximations of the statistical properties of the problem are also necessary if the method is not to be too computationally intensive. Provided that these properties are recognized, DLT can be used successfully in work with non-metric cameras that does not require rigorous design and estimates of data quality.

2.4 Multistation convergent geometry

In Figure 2.13, five cameras are shown disposed around an object having a typical target at A_i which gives rise to an image point a_{ij} in camera j. In general, each target A_i is imaged in each camera, but not necessarily. The collinearity equations (2.16) for a_{ij} are

$$x_{ij} = \frac{-c_j\left[r_{j,11}\left(X_{O_j} - X_i\right) + r_{j,12}\left(Y_{O_j} - Y_i\right) + r_{j,13}\left(Z_{O_j} - Z_i\right)\right]}{\left[r_{j,31}\left(X_{O_j} - X_i\right) + r_{j,32}\left(Y_{O_j} - Y_i\right) + r_{j,33}\left(Z_{O_j} - Z_i\right)\right]} \qquad (2.28a)$$

$$y_{ij} = \frac{-c_j\left[r_{j,21}\left(X_{O_j} - X_i\right) + r_{j,22}\left(Y_{O_j} - Y_i\right) + r_{j,23}\left(Z_{O_j} - Z_i\right)\right]}{\left[r_{j,31}\left(X_{O_j} - X_i\right) + r_{j,32}\left(Y_{O_j} - Y_i\right) + r_{j,33}\left(Z_{O_j} - Z_i\right)\right]} \qquad (2.28b)$$

for negative images. For positive images, the principle distance c_j has a positive sign. If differential image scale is suspected, but has not been corrected during image refine-

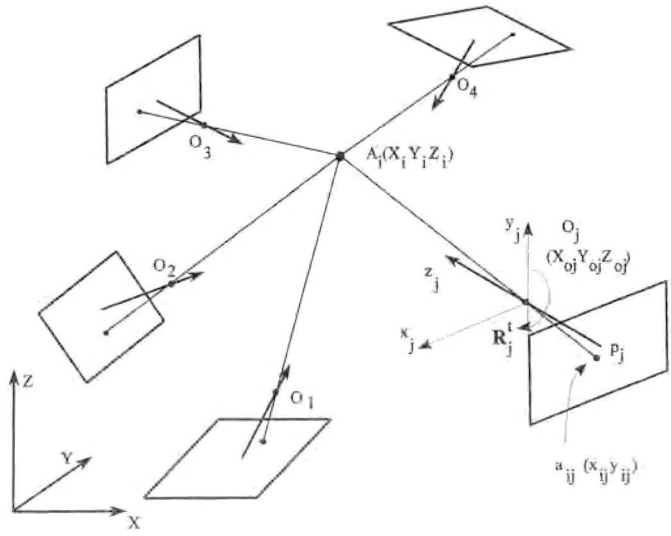

Figure 2.13 *Multistation convergent configuration.*

ment, two different principal distances c_{jx} and c_{jy} could be used in (2.28a) and (2.28b) respectively. If more additional parameters are to be included, a typical set for non-metric cameras could comprise, for camera j: two interior orientation elements (x_{oj}, y_{oj}); coefficients K_{1j}, K_{2j} and K_{3j} of the polynomial (2.17) representing geometrical radial lens distortion; and coefficients P_{1j} and P_{2j} of the polynomial (2.18) representing the geometrical effects of lens decentring. Then the left-hand sides of the collinearity equations (2.28a) and (2.28b) will become

$$x_{ij} - x_{oj} + (x_{ij} - x_{oj})\, r_{ij}^{-1} (K_{1j} r_{ij}^3 + K_{2j} r_{ij}^5 + K_{3j} r_{ij}^7) + P_{1j}[r_{ij}^2 + 2(x_{ij} - x_{oj})^2] + 2P_{2j}(x_{ij} - x_{oj})(y_{ij} - y_{oj}) \quad (2.29a)$$

and

$$y_{ij} - y_{oj} + (y_{ij} - y_{oj})\, r_{ij}^{-1} (K_{1j} r_{ij}^3 + K_{2j} r_{ij}^5 + K_{3j} r_{ij}^7) + P_{2j}[r_{ij}^2 + 2(y_{ij} - y_{oj})^2] + 2P_{1j}(x_{ij} - x_{oj})(y_{ij} - y_{oj}) \quad (2.29b)$$

respectively, where $r_{ij}^2 = (x_{ij} - x_{oj})^2 + (y_{ij} - y_{oj})^2$. Equations (2.28a) - (2.29b) inclusive can be written

$$\mathbf{F}(\mathbf{x}, \mathbf{b}, \mathbf{a}) = \mathbf{0} \quad (2.30)$$

where $\mathbf{x}$ is a vector representing the u elements whose values are required (the *parameters to be estimated*) $\mathbf{b}$ is a vector representing the m measured elements and $\mathbf{a}$ is a vector representing elements whose values are known constants. Equation (2.30) is a *functional*

model of the photogrammetry based on collinearity equations. It is used to evaluate **x**, given values of **b** and **a**. Generally no unique solution for **x** exists, but least squares estimation provides a means for obtaining one and is discussed in the next section. Choices often must be made as to which elements are measurements, which are parameters to be estimated and which constants. The vector **b** includes all photo co-ordinates x_{ij} and y_{ij} and any measurements made on or around the object, such as the distance s_{pq} between targets P and Q. The functional model (2.30) will then include the additional equation

$$s_{pq} - [(X_q - X_p)^2 + (Y_q - Y_p)^2 + (Z_q - Z_p)^2]^{1/2} = 0 \qquad (2.31)$$

which can be extended by including elements related to the calibration of the instrument used for the measurement of the distance s_{pq}. Such elements could be a zero correction Δ and a scale factor $(1 + \mu)$ leading to

$$s_{pq}(1 + \mu) + \Delta - [(X_q - X_p)^2 + (Y_q - Y_p)^2 + (Z_q - Z_p)^2]^{1/2} = 0 \qquad (2.32)$$

The co-ordinates X_i, Y_i and Z_i of the targets will generally be included in **x**. If the cameras have undergone prior calibration and accurate, reliable values obtained for principal point offsets, principal distances and lens distortion coefficients for example (*see* Chapter 6) these calibrated values could be included in **a**. An alternative procedure with prior camera calibration data is to use the prior values of the parameters in **b** as measurements (and make use of their standard deviations) but include the same parameters also in **x**. If no prior calibration values have been obtained, it is possible to include calibration elements in **x** only. This procedure is referred to as *self-calibration*. Similar options exist for other calibration elements such as Δ and μ (but if μ is included in **x** as an unknown parameter to be estimated, datum problems may arise - see the next section) and for other elements of the functional model. For example, the elements of exterior orientation of the cameras may have been evaluated by a prior process so if it is reasonable to assume they have remained unchanged, the values can be included in the next process as either constants (in **a**) or as measurements (in **b** and in **x**).

2.5 Least squares estimation (LSE)

Although it is possible to justify the use of LSE solely on the basis of statistical probabilities, that is not the main reason why it is widely used in geodesy, engineering surveying, close range photogrammetry and other disciplines where more (often many more) measurements are available than the minimum necessary to evaluate the unknown elements. LSE provides a systematic method for computing unique values of co-ordinates and other elements in close range photogrammetry based on a large number of redundant measurements of different kinds and weights. It allows for covariance matrices of estimates to be readily derived from the covariance matrix of the measurements. If a covariance matrix of the measurements is assumed, a priori analysis can be used to design a camera/object configuration and measurement scheme to meet criteria relating to precision, reliability and accuracy. This attribute of LSE is particularly useful

in close range photogrammetry where almost every measurement task has unique features. LSE is also flexible: it allows elements to be treated as unknowns, or as measurements, or as constants depending on circumstances; and algorithms within a LSE process can be devised to suit particular measurement tasks. The main disadvantage of LSE is that it does not make it easy for the photogrammetrist to identify blunders in measurements. The non-linear nature of the functional model (2.30) means that for practical reasons approximate values of the unknowns must first be determined so that iterated linear solutions can be carried out in preference to a non-linear LSE process. Despite these difficulties, the general usefulness of linearized LSE processes has been demonstrated so often that an alternative is not urgently sought when rigorous statistical analysis is necessary.

The functional model is given by the set of c equations represented by (2.30) where **x** represents the u unknown elements to be estimated and **b** represents the m measured elements. Constants represented by **a** in (2.30) are usually implied, so that the functional model is then

$$\mathbf{F}(\mathbf{x}, \mathbf{b}) = 0. \tag{2.33}$$

It is customary to assume measurements are independent random variables, each measurement b_i having standard deviation σ_i and *weight* w_i given by

$$w_i = \sigma_o^2 / \sigma_i^2 \tag{2.34}$$

where σ_o^2 is the *variance factor*, or *reference variance* or *standard deviation of a measurement of unit weight*. The *weight matrix* of the m measurements is the $m \times m$ diagonal matrix $\mathbf{W}$ whose non-zero elements are the weights w_i. The covariance matrix of independent measurements is the $m \times m$ diagonal matrix $\mathbf{C}_b$ whose non-zero elements are the variances σ_i^2. A cofactor matrix **Q** is defined by

$$\mathbf{Q} = \sigma_o^{-2}\mathbf{C} \tag{2.35}$$

so that for independent measured elements **b**

$$\mathbf{Q}_b = \sigma_o^{-2}\mathbf{C}_b = \mathbf{W}^{-1} \tag{2.36}$$

An a priori value $\sigma_o^2 = 1$ can be assigned. In this special case $\mathbf{W} = \mathbf{C}_b^{-1} = \mathbf{Q}_b^{-1}$. The numerical requirement is: given measured values of the m elements **b**, their weight matrix $\mathbf{W}$ and the c equations in the functional model $\mathbf{F}(\mathbf{x}, \mathbf{b}) = 0$, evaluate the u parameters **x** and their covariance matrix $\mathbf{C}_x$. This is not straightforward because the functional model is not linear, m and c are usually greater than u (there is a redundancy of data), and the measured values are subject to random errors (the data are inconsistent). The difficulty of having redundant and inconsistent numerical data is overcome by using the principle of weighted least squares to obtain a unique solution. The first difficulty is overcome by linearizing the functional model about approximate values of the random variables (measured variables and those to be estimated) and finding the unique

solution by iteration. The computational process should be followed by analysis of the solution and its associated statistics to define the quality of the measurements and least squares estimates. Conformance to a specification can then be assessed.

Taylor's theorem can be used to linearize the functional model (2.33) by partial differentiation with respect to measurements **b** and parameters to be estimated **x**:

$$\mathbf{F}(\mathbf{x},\mathbf{b}) = \mathbf{F}(\mathbf{x}_o,\mathbf{b}_o) + \left(\frac{\delta \mathbf{F}}{\delta \mathbf{x}}\right)_o (\mathbf{x}-\mathbf{x}_o) + \left(\frac{\delta \mathbf{F}}{\delta \mathbf{b}}\right)_o (\mathbf{b}-\mathbf{b}_o) \tag{2.37}$$

to first order accuracy only or, numerically

$$\mathbf{F}(\bar{\mathbf{x}},\bar{\mathbf{b}}) = \mathbf{F}(\mathbf{x}_o,\mathbf{b}_o) + \mathbf{A}(\bar{\mathbf{x}}-\mathbf{x}_o) + \mathbf{D}(\bar{\mathbf{b}}-\mathbf{b}_o) = 0 \tag{2.38}$$

where $\bar{\mathbf{x}}$ and $\bar{\mathbf{b}}$ represent values that fit the functional model exactly (true values) and $\mathbf{x}_o$ and $\mathbf{b}_o$ are first order approximations to those values. In practice, measured values can be used for $\mathbf{b}_o$. The term $(\bar{\mathbf{b}} - \mathbf{b}_o) = \mathbf{v}$ then represents corrections to measurements necessary to satisfy the functional model exactly. These corrections are numerically the same as the *measurement residuals*, but of opposite sign. Therefore (2.38) can be written

$$\mathbf{A}(\bar{\mathbf{x}} - \mathbf{x}_o) + \mathbf{D}\mathbf{v} + \mathbf{F}(\mathbf{x}_o, \mathbf{b}_o) = 0 \tag{2.39}$$

or

$$\mathbf{A}\mathbf{x} + \mathbf{D}\mathbf{v} = \mathbf{d} \tag{2.40}$$

where $\mathbf{x} = (\bar{\mathbf{x}} - \mathbf{x}_o)$ and $\mathbf{d} = -\mathbf{F}(\mathbf{x}_o, \mathbf{b}_o)$. The unique *least squares estimates* of $\bar{\mathbf{x}}$ and $\bar{\mathbf{b}}$ (denoted by $\hat{\mathbf{x}}$ and $\hat{\mathbf{b}}$) are those that satisfy the weighted least squares criterion

$$\varphi = \mathbf{v}^t\mathbf{W}\mathbf{v} \rightarrow \text{minimum}$$

or

$$\varphi = \mathbf{v}^t\mathbf{W}\mathbf{v} + 2\mathbf{k}^t(\mathbf{A}\mathbf{x} + \mathbf{D}\mathbf{v} - \mathbf{d}) \rightarrow \text{minimum}$$

where **k** is a $c \times 1$ vector of *Lagrangian multipliers* introduced so that the estimates of **x** and **v** can be found from

$$\begin{bmatrix} \mathbf{W} & \mathbf{D}^t & \mathbf{0} \\ \mathbf{D} & \mathbf{0} & \mathbf{A} \\ \mathbf{0} & \mathbf{A}^t & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{k} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{d} \\ \mathbf{0} \end{bmatrix}. \tag{2.41}$$

These estimates are:

$$\hat{\mathbf{x}} = [\mathbf{A}^t(\mathbf{D}\mathbf{W}^{-1}\mathbf{D}^t)^{-1}\mathbf{A}]^{-1}\mathbf{A}^t(\mathbf{D}\mathbf{W}^{-1}\mathbf{D}^t)^{-1}\mathbf{d} \tag{2.42}$$

and

$$\hat{\mathbf{v}} = \mathbf{W}^{-1}\mathbf{D}^t(\mathbf{D}\mathbf{W}^{-1}\mathbf{D}^t)^{-1}\{\mathbf{I} - \mathbf{A}[\mathbf{A}^t(\mathbf{D}\mathbf{W}^{-1}\mathbf{D}^t)^{-1}\mathbf{A}]^{-1}\mathbf{A}^t(\mathbf{D}\mathbf{W}^{-1}\mathbf{D}^t)^{-1}\}\mathbf{d} \tag{2.43}$$

from which the estimated measurements $\hat{\mathbf{b}} = \mathbf{b} + \hat{\mathbf{v}}$.

It is often possible in practice to arrange for all necessary functional relationships between measured and unknown elements to be explicit in the measured elements. The collinearity equations (2.16a) and (2.16b) are examples. This procedure greatly simplifies least squares processes and makes them more flexible. Equation (2.33) then becomes

$$\mathbf{F}(\mathbf{x}, \mathbf{b}) \equiv \mathbf{f}(\mathbf{x}) - \mathbf{b} = \mathbf{0} \tag{2.44}$$

and (2.39) reduces to

$$\mathbf{A}\mathbf{x} = \mathbf{b}_o - \mathbf{f}(\mathbf{x}_o) + \mathbf{v} \tag{2.45}$$

by replacing $(\bar{\mathbf{x}} - \mathbf{x}_o)$ by $\mathbf{x}$ and noting that $\mathbf{D} = -\mathbf{I}$, the $m \times m$ identity matrix. Equations (2.38) to (2.43) deal with the *general linearized least squares problem* and (2.45) together with $\mathbf{W}$ is a special case, referred to in statistics as the *Gauss-Markov linear model* and often by photogrammetrists as *linearized observation equations*. $\mathbf{A}$ is the $m \times u$ matrix of partial differential coefficients $(\partial \mathbf{f}/\partial \mathbf{x})_o$ evaluated at $\mathbf{x} = \mathbf{x}_o$, $\mathbf{b}_o$ is the $m \times 1$ vector of measured values, $\mathbf{f}(\mathbf{x}_o)$ is the $m \times 1$ vector of values of the measured elements computed using $\mathbf{x}_o$ (so $\mathbf{b}_o - \mathbf{f}(\mathbf{x}_o) = \mathbf{b}$ is often referred to as the *observed - computed term*) and $\mathbf{v}$ is the $m \times 1$ vector of least squares corrections whose numerical values must satisfy the requirement that $\varphi = \mathbf{v}^t\mathbf{W}\mathbf{v}$ is a minimum. This condition is satisfied by $\hat{\mathbf{x}}$ given by the *u normal equations*

$$\mathbf{A}^t\mathbf{W}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^t\mathbf{W}\mathbf{b} \tag{2.46}$$

where $\hat{\mathbf{x}}$ is a first order increment to $\mathbf{x}_o$ to give the values of the least squares estimates of the u unknowns $\mathbf{x}$. This process of evaluating co-ordinates of targets and exterior orientation parameters of the cameras using least squares based on collinearity equations is often referred to by photogrammetrists as a *bundle adjustment*. When parameters that define the camera calibration are also evaluated, the process is called a *self-calibrating bundle adjustment*.

If $\mathbf{A}^t\mathbf{W}\mathbf{A}$ is of full rank

$$\hat{\mathbf{x}} = (\mathbf{A}^t\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^t\mathbf{W}\mathbf{b} \tag{2.47}$$

and assuming $\sigma_o^2 = 1$

$$\mathbf{C}_{\hat{x}} = (\mathbf{A}^t\mathbf{W}\mathbf{A})^{-1} . \tag{2.48}$$

2.5.1 Definition of the co-ordinate datum in object space

The solution for $\hat{\mathbf{x}}$ cannot in general be obtained from computing the regular (Cayley) matrix inverse of $\mathbf{A}^t\mathbf{W}\mathbf{A}$ as shown in (2.47) and (2.48) because the matrix is singular unless procedures are adopted to remove its rank deficiency. Rank deficiencies are of two kinds: defects arising from the camera/object configuration and measured elements,

and defects arising from an incomplete definition of the elements of the co-ordinate datum in object space. An illustration of a cause of configuration defect is a target whose spatial co-ordinates (for which no prior values are available) are included in $\mathbf{x}$, and for which the only measurements are the (x,y) photoco-ordinates of its image on one photograph. Configuration defects lead to linear dependence between the rows of $\mathbf{A}$ and hence of $\mathbf{A}^t\mathbf{W}\mathbf{A}$. They must be avoided by proper design of the configuration and careful checking, preferrably automated, of the measurement scheme.

Datum defects, on the other hand, give rise to linear dependence between the columns of $\mathbf{A}$ (and between the columns and rows of $\mathbf{A}^t\mathbf{W}\mathbf{A}$). An object space co-ordinate system for positioning in three dimensions needs seven datum elements to be defined: one for scale (unless it is anisotropic), three for translation (origin) and three for rotation (directions of axes). In a purely photogrammetric configuration (i.e. one in which the only measurements are photo co-ordinates of images of targets) in which no prior numerical information about the locations of targets or cameras is included in the functional model, none of the datum elements is defined so $\mathbf{A}^t\mathbf{W}\mathbf{A}$ has a column rank defect of seven. If measurements include one or more slope distances the contribution of one or more equations such as (2.31) will remove one deficiency by defining the object scale (but an equation such as (2.32) will not define scale because it includes the scale parameter μ). If a vertical distance (height difference) or a horizontal distance in the object space is measured, four datum elements of a Cartesian system will be left undefined: the three elements of translation and the direction of one of the axes in the horizontal plane. Once this direction is defined, the direction of the third axis will also be defined because the Cartesian system is right-handed and orthogonal. Although measurements in object space might be used to reduce the rank deficiency of $\mathbf{A}^t\mathbf{W}\mathbf{A}$ (or to increase reliability – *see* section 2.6) they can seldom be made to sufficient precision to contribute significantly to the overall precision of estimated positions. In many cases in close range photogrammetry, photogrammetric measurements (i.e. photoco-ordinates of images of targets) alone give much higher precision of object space co-ordinates than is possible when only direct measurements are used. On the other hand in small and medium scale aerial mapping, ground control should always increase the overall precision of photogrammetry significantly. To what extent, and why, object space measurements should be used in close range photogrammetry is a matter to be decided at the design stage.

2.5.2 Definition of co-ordinate datum elements by minimum constraint equations

When it is not possible, or very difficult, to devise measurements of adequate precision in the object space, datum defects can be removed by analytical methods. For a purely photogrammetric network, the seven datum defects can be rectified by the inclusion in the functional model of seven *constraint equations* of the correct type. Constraint equations have the form $\mathbf{g}(\mathbf{x}) = \mathbf{a}$, where $\mathbf{a}$ is a constant vector. If some of the seven datum elements are defined by measurements included in the functional model as observation equations, only those constraint equations necessary to complete the datum definition must be included. The network is then a *minimum constraints network*. The number d of minimum constraints needed to remove the rank defect in $\mathbf{A}$ will be seven or less.

When the constraint equations $\mathbf{g(x)} = \mathbf{a}$ are linearized using Taylor's theorem they can be written as $\mathbf{Bx} = \mathbf{0}$ and included with the linearized observation equations to give

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{b} + \mathbf{v} \\ \mathbf{0} \end{bmatrix} . \tag{2.49}$$

The least squares estimate of $\mathbf{x}$ is found by minimizing $\varphi = \mathbf{v}^t\mathbf{W}\mathbf{v}$ and is given by the *augmented normal equations*

$$\begin{bmatrix} \mathbf{A}^t\mathbf{W}\mathbf{A} & \mathbf{B}^t \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{k} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^t\mathbf{W}\mathbf{b} \\ \mathbf{0} \end{bmatrix} \tag{2.50}$$

where $\mathbf{k}$ is the $d \times 1$ vector of *Lagrangian multipliers*, d being the number of minimum constraint equations. The $u \times u$ singular matrix $\mathbf{A}^t\mathbf{W}\mathbf{A}$ having a rank deficiency d is augmented by the addition of d rows and d columns to give a $[u+d] \times [u+d]$ non-singular matrix. Although this matrix is symmetric it is not positive definite (because of the d zero elements on the leading diagonal) so routines for computing the inverse of a positive definite symmetric matrix (such as $\mathbf{A}^t\mathbf{W}\mathbf{A}$ when of full rank) cannot be used.

Although the number d and type of minimum constraint equations are determined by the number and type of datum defects, there is a wide choice of equations for implementation. This can be illustrated by a simple but common example in close range photogrammetry. Usually the three positional elements of the datum are undefined by measurements so three constraint equations are necessary. These equations could be

$$X_i = a_1 ; \qquad Y_i = a_2 ; \quad \text{and} \quad Z_i = a_3 \tag{2.51}$$

where i is any point (target or camera perspective centre for example) in the configuration (in fact, the three co-ordinate constraints need not all be applied to the same point i) and a_1, a_2, etc. are given constant values. Linearization of these equations gives

$$\begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \delta X_1 & \cdots & \delta Z_{i-1} & \delta X_i & \delta Y_i & \delta Z_i & \delta X_{i+1} & \cdots & \delta Z_p \end{bmatrix}^t = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

assuming p points whose object space co-ordinates are to be estimated. Similar linearized constraint equations can be derived from:

$$[(X_j - X_i)^2 + (Y_j - Y_i)^2 + (Z_j - Z_i)^2]^{1/2} = a_4$$

for scale; and

$$\arctan[(X_j - X_i)/(Y_j - Y_i)] = a_5$$

for rotation (about the Z axis in this case). The choice of minimum constraint equations will have an influence on the values of the elements of $\hat{\mathbf{x}}$ and on the values of the elements of its covariance matrix $\mathbf{C}_{\hat{x}}$ because $\hat{\mathbf{x}}$ and $\mathbf{C}_{\hat{x}}$ are not *datum invariant*. Referring to (2.50), let

$$\begin{bmatrix} \mathbf{A}^t\mathbf{W}\mathbf{A} & \mathbf{B}^t \\ \mathbf{B} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{N}_{11} & \mathbf{N}_{12} \\ \mathbf{N}_{12}^t & \mathbf{0} \end{bmatrix} \tag{2.52}$$

so the least squares estimate can be written as $\hat{\mathbf{x}} = \mathbf{N}_{11}\mathbf{A}^t\mathbf{W}\mathbf{b}$ and its (singular) covariance matrix is $\mathbf{C}_{\hat{x}} = \mathbf{N}_{11}$ assuming $\sigma_o^2 = 1$. The influence of the choice of constraint equations on $\mathbf{C}_{\hat{x}}$ can be seen by considering the constraints defined by (2.51). They produce zero values for all elements in the rows and columns of $\mathbf{C}_{\hat{x}}$ that relate to the X, Y and Z co-ordinates of point i; all the other elements in $\mathbf{C}_{\hat{x}}$ are non-zero. In such a case, point i is said to be part of the *zero variance reference base* of the network. The choice of constraints is therefore quite important because it will define the numerical values of the estimated co-ordinates and influence their variances and covariances. For a configuration with a rank defect of seven, constraint equations can be applied to the X, Y and Z co-ordinates of any two points and to the X, or Y, or Z co-ordinate of a third (non-collinear) point. The zero variance reference base is then fully defined by those seven co-ordinates. Their estimated co-ordinates will be the values assigned to them by the constraint equations. The variances and covariances relating to those co-ordinates will all be zero; the co-ordinates are not random variables, but fixed values. If the d minimum constraints are to be applied solely by fixing d co-ordinates, it is possible to remove those co-ordinates from $\mathbf{x}$, the vector of unknowns to be estimated, to give $\hat{\mathbf{x}}$ and remove the corresponding d columns in $\mathbf{A}$ that operate on them to form another matrix $\mathbf{A}_1$ with full column rank. The regular Cayley inverse $(\mathbf{A}_1^t\mathbf{W}\mathbf{A}_1)^{-1}$ can then be computed and used for the least squares estimation of $\hat{\mathbf{x}}_1$ as in (2.47) and (2.48). This procedure is easier than using linearized constraint equations, but not as flexible. It may be necessary subsequently to expand $(\mathbf{A}_1^t\mathbf{W}\mathbf{A}_1)^{-1}$ to give $\mathbf{C}_{\hat{x}}$ (which will be singular) by placing d rows and columns of zero elements in the appropriate locations and to expand $\hat{\mathbf{x}}_1$ by placing (in corresponding locations) the d fixed co-ordinate values to give $\hat{\mathbf{x}}$.

Given the almost limitless possibilities for forming minimum constraint equations, it is useful to consider a unique set of minimum constraint equations that in a specific sense is 'better' than any other. It is 'best' in the sense that it gives rise to a covariance matrix $\mathbf{C}_{\hat{x}}$ that has the minimum trace of all possible covariance matrices. Such a set of constraints is usually referred to as *inner constraints*. Its use results in the minimum sum of variances of the estimated co-ordinates of the points to which the constraints are applied. For a configuration defect of seven, the full set of linearized constraint equations is as follows. For the three positional constraints:

$$\Sigma\delta X_i = \Sigma\delta Y_i = \Sigma\delta Z_i = 0 .$$

For the three rotational constraints:

$$\Sigma[Z_i \delta Y_i - Y_i \delta Z_i] = \Sigma[-Z_i \delta X_i + X_i \delta Z_i] = \Sigma[Y_i \delta X_i - X_i \delta Y_i] = 0,$$

and for the scale constraint:

$$\Sigma[X_i \delta X_i + Y_i \delta Y_i + Z_i \delta Z_i] = 0.$$

The summations are over the range $i = 1$ to p corresponding to the set of p points for which the sum of variances of estimated co-ordinates is to be minimized. The set usually comprises all targetted points, but sometimes a specific subset of them is selected. It is not usual to apply inner constraints to camera parameters. The form of the linearized inner constraint equations is

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & .... & 0 \\ 0 & 1 & 0 & 0 & 1 & .... & 0 \\ 0 & 0 & 1 & 0 & 0 & .... & 1 \\ 0 & Z_1 & -Y_1 & 0 & Z_2 & .... & -Y_p \\ -Z_1 & 0 & X_1 & -Z_2 & 0 & .... & X_p \\ Y_1 & -X_1 & 0 & Y_2 & -X_2 & .... & 0 \\ X_1 & Y_1 & Z_1 & X_2 & Y_2 & .... & Z_p \end{bmatrix} \begin{bmatrix} \delta X_1 \\ \delta Y_1 \\ \delta Z_1 \\ \delta X_2 \\ \delta Y_2 \\ .... \\ .... \\ \delta Z_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (2.53)$$

Denoting these constraint equations by $\mathbf{Gx} = 0$, the normal equations will be of the form in (2.52) but with G instead of **B**. The sub-matrix corresponding to $\mathbf{N}_{11}$ in (2.52) will be singular, but its trace will be the minimum of all traces of $p \times p$ covariance matrices relating to the p points derived by minimum constraints, including those with a zero variance reference base.

2.5.3 Definition of co-ordinate datum elements by stochastic constraints

Other ways of defining elements of the co-ordinate datum are possible and sometimes more appropriate than using minimum constraint equations. If a self-calibrating bundle adjustment for an array of cameras has been carried out using a calibration field, and it can be reasonably assumed that exterior and interior orientation elements and calibration parameters of the cameras have not changed since then, their estimated values and covariance matrix can be incorporated in subsequent LSE processes relating to photogrammetry of another object. Suppose the previously estimated elements are denoted by $\mathbf{x}_p$ and their prior values by $\mathbf{x}_p^\circ$ with covariance matrix $\mathbf{C}_{xp}$. The new linearized observation equations $\mathbf{Ax} = \mathbf{b} + \mathbf{v}$ can be written by partitioning **x** into $\mathbf{x}_n$ (representing the new parameters to be estimated, which in the case cited will consist solely of co-ordinates of targets in object space) and $\mathbf{x}_p$, the previously estimated elements (but not their values):

$$\begin{bmatrix} \mathbf{A}_n & \mathbf{A}_p \end{bmatrix} \begin{bmatrix} \mathbf{x}_n \\ \mathbf{x}_p \end{bmatrix} = \mathbf{b}. \tag{2.54}$$

In general **A** will be singular with a rank defect of 7. If however the previously estimated values $\mathbf{x}_p^\circ$ are included as measurements, the augmented observation equations become

$$\begin{bmatrix} \mathbf{A}_n & \mathbf{A}_p \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}_n \\ \mathbf{x}_p \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{x}_p^\circ \end{bmatrix} \tag{2.55}$$

with cofactor matrix

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{xp} \end{bmatrix} \tag{2.56}$$

to give the normal equations

$$\begin{bmatrix} \mathbf{A}_n^t\mathbf{Q}_b^{-1}\mathbf{A}_n & \mathbf{A}_n^t\mathbf{Q}_b^{-1}\mathbf{A}_p \\ \mathbf{A}_p^t\mathbf{Q}_b^{-1}\mathbf{A}_n & \mathbf{A}_p^t\mathbf{Q}_b^{-1}\mathbf{A}_p + \mathbf{Q}_{xp}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_n \\ \mathbf{x}_p \end{bmatrix} = \begin{bmatrix} \mathbf{A}_n^t\mathbf{Q}_b^{-1}\mathbf{b} \\ \mathbf{A}_p^t\mathbf{Q}_b^{-1}\mathbf{b} + \mathbf{Q}_{xp}^{-1}\mathbf{x}_p^\circ \end{bmatrix}. \tag{2.57}$$

The least squares estimates $\hat{\mathbf{x}}_n$ and $\hat{\mathbf{x}}_p$ can be evaluated using the regular inverse of the coefficient matrix of the normal equations. If the only measurements in **b** were photo co-ordinates, the matrix $\mathbf{A}^t\mathbf{W}\mathbf{A}$, or

$$\begin{bmatrix} \mathbf{A}_n^t\mathbf{Q}_b^{-1}\mathbf{A}_n & \mathbf{A}_n^t\mathbf{Q}_b^{-1}\mathbf{A}_p \\ \mathbf{A}_p^t\mathbf{Q}_b^{-1}\mathbf{A}_n & \mathbf{A}_p^t\mathbf{Q}_b^{-1}\mathbf{A}_p \end{bmatrix}$$

would have a rank deficiency of 7. The addition of $\mathbf{Q}_{xp}^{-1}$ to the lower right sub-matrix of $\mathbf{A}^t\mathbf{W}\mathbf{A}$ removes its rank deficiency and the regular matrix inverse can be computed. It is important to recall that the computations are iterative; in (2.55) and (2.57) the elements in $\mathbf{x}_n$ and $\mathbf{x}_p$ are in fact first order increments to approximate values of the parameters to be estimated and that **b** and $\mathbf{x}_p^\circ$ are measured values minus the values of the measured elements computed from the approximate values of the parameters to be estimated. When $\mathbf{x}_p^\circ$ represents prior values of exterior orientation elements and calibration parameters of the cameras, those values serve as approximate values, so for the first iteration the lower sub-vector on the right-hand side of equation (2.55) will be null. Subsequent iterations will cause changes to the prior values (and their cofactor matrix) as a result of the new measurements and LSE processes. The use of prior values of parameters with their cofactor matrix is therefore clearly different from the use of constraint equations such as those in (2.51). However the equations $\mathbf{x}_p = \mathbf{x}_p^\circ + \mathbf{v}_p$ in (2.55) with $\mathbf{Q}_{xp}$ can be regarded as *stochastic constraints* in contrast to equations such as (2.49) which include *deterministic constraints*.

2.5.4 *Some non-rigorous procedures*

The foregoing LSE procedures are rigorous, but the time needed to carry out some of the computations may restrict their usefulness in some cases. Simpler, faster, but non-rigorous computational procedures for estimating object space co-ordinates of targets can be used by regarding prior values of exterior orientation elements (from resection for example, section 2.2.4) and calibration parameters of the cameras as fixed. Then in the collinearity equations (2.16 a and b) supplemented by (2.18 a and b) if necessary, all elements except X_i, Y_i and Z_i (unknowns to be estimated) and x_{ij} and y_{ij} (measurements) are regarded as known constants. Linearized observation equations can be set up to evaluate each X_i, Y_i and Z_i in turn by intersection as described in section 2.3, each time necessitating the computation of the inverse of only a 3 × 3 matrix. This procedure is much faster than using the prior values as stochastic constraints, when they are re-evaluated at the same time as the co-ordinates of targets; computing the inverses of say 100 3×3 matrices is faster than computing the inverse of one matrix with dimensions equal to 300 plus the number of camera parameters.

2.6 Data quality

Automation of image acquisition and measurement is a feature of close range digital photogrammetry that can lead to its incorporation in many technological processes as a measurement system. Automation continues as measured data (photo co-ordinates, in general) are transformed into three dimensional co-ordinates in the object space by LSE processes described in the preceding section. On the other hand, when photographs on glass or film are measured by an operator using a comparator, the photogrammetrist is able to monitor or screen the data fairly closely and take the opportunity to correct or remeasure data that are clearly in error. It is arguable whether errors introduced by human fallibility in measurement outweigh errors that can be detected and rectified by human intelligence. When automated processes largely replace direct human activity, automated methods of data screening must be part of the automated processes; in other words, the quality of the data that are automatically generated must be automatically assessed. The fact that LSE processes are the basis of computation in close range photogrammetry means that several indicators of the quality of measured and derived numerical data are available, not only to assess the quality of data after the measurement has been completed, but also to design a measuring system to meet specific requirements (Chapter 9).

Data quality can be defined in relation to accuracy, precision and reliability. *Accuracy* can be increased by calibrating measuring equipment and by applying corresponding corrections to measured values or by improving the functional model (2.30) that defines the mathematical relationship between measured and derived elements. For example, the accuracy of co-ordinates such as (X_i, Y_i, Z_i) derived from (2.28a) and (2.28b) as a functional model can be increased either by replacing the left-hand sides of those equations by the expressions in (2.29a) and (2.29b) to take account of the effects of lens distortions, or by calibrating the cameras independently and computing and applying corrections to photo co-ordinates before starting the LSE process. Whether the result-

ant increase in accuracy will be significant or not should be decided at the design stage. If the effects of lens distortions defined in (2.29a) and (2.29b) are ignored, *systematic errors* will be present. Similarly, systematic errors will arise if (2.31) is used instead of (2.32); accuracy can be improved either by calibration of the measuring equipment for scale and zero error and correction of the measured value accordingly, or by including μ and Δ as additional parameters to be evaluated by LSE.

Precision is related to the random errors in measurements and derived values, and is characterized by covariance matrices. Although it is not always necessary to make assumptions about the form of the probability density function for a particular random variable it is usually assumed that measured values are independent and uncorrelated. Values of elements derived by LSE on the other hand are generally correlated through the LSE processes that are used to evaluate them. For any random variable, measured or evaluated by LSE, its variance (or standard deviation) is an indication of its precision; the smaller the variance the higher the precision. The precision of the mean value of a measurement will always be improved by increasing the number of measurements.

Reliability is related to gross errors in measurements. Unlike systematic errors, the effects of which can be removed (or greatly reduced) by applying corrections or by including additional parameters in the LSE processes, and random errors which can be characterized and assessed using statistical analysis, gross errors arise through mistakes in procedure, or *blunders*. The detection of gross errors in the context of LSE is difficult because of the target function ($\varphi = \mathbf{v}^t\mathbf{W}\mathbf{v}$) that is minimised: if one measurement has a gross error, instead of that measurement having a correspondingly large residual or correction after LSE, the error will be spread amongst many measurements in order to minimize φ. It might then be difficult to detect that a gross error were present and even more difficult to identify the measurement that has the error. In order to detect the presence or absence of a gross error in a measurement, it is necessary to make some assumptions about the probability density function of the measurement because the detection procedure is a statistical test of hypotheses. For a given confidence level (usually 95%) the smaller the gross error that can be detected in a given measurement, the higher the reliability of the photogrammetric process. Such reliability is referred to as *internal reliability*. *External reliability* on the other hand is the effect that an undetected gross error in a measurement has on the estimated parameters. An object/camera configuration has a high internal reliability if it permits the detection of small gross errors and a high external reliability if the presence of one or more large gross errors has little effect on the values of the estimated parameters.

Many of the statistics that are used for assessing data quality are found in covariance matrices of measured and derived values of elements of the functional model. Assuming that $\sigma_o^2 = 1$, the covariance matrix $\mathbf{C}_{\hat{x}}$ of the estimated parameters $\hat{\mathbf{x}}$ is given by (2.48) if $\mathbf{A}$ is of full rank and by $\mathbf{N}_{11}$ in (2.52) if it is not and minimum constraints are used. The covariance matrix $\mathbf{C}_{\hat{v}}$ of the least squares estimate of the corrections $\hat{\mathbf{v}}$ ($= \mathbf{f}(\hat{\mathbf{x}}) - \mathbf{b}$) from (2.44) to the measurements $\mathbf{b}$ is

$$\mathbf{C}_{\hat{v}} = \mathbf{W}^{-1} - \mathbf{A}(\mathbf{A}^t\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^t \qquad (2.58)$$

and the covariance matrix $\mathbf{C}$ of the corrected measurements $\hat{\mathbf{b}}$ ($= \mathbf{b} + \hat{\mathbf{v}}$) is

$$C_{\hat{l}} = W^{-1} - C_{\hat{v}} = A(A^tWA)^{-1}A^t \tag{2.59}$$

again assuming that the variance factor is unity.

2.6.1 Tests using all the least squares corrections (or residuals)

Some of the indicators of data quality rely on the least squares corrections having a normal distribution with expectation zero. A visual display of the magnitudes and signs of the corrections allows the photogrammetrist to see whether any gross errors are obviously present and ought to be located and removed. A χ^2 goodness-of-fit test can be carried out to see whether or not the corrections are normally distributed with expectation zero. Often a few hundred photoco-ordinates are measured, sometimes more, so a reasonable class interval can usually be chosen as the basis for the test. Another useful indication of the characteristics of the corrections is a display, in the format of each camera image, of the photo co-ordinate correction vectors at a suitable scale; any non-random pattern that can be seen in them indicates the likely presence of systematic errors, probably related to the calibration of that particular camera.

The third order design (Chapter 9) is to decide on the standard deviations of the measurements. When the measurements have been made and are to be processed by LSE it is necessary to define their weights. This is usually done by (2.34) but with σ_o^2 put equal to unity a priori. This implies that each measured value has its designed standard deviation σ_i^2. After LSE it is necessary to find out whether or not the measurements had been made to their designed precisions. An a posteriori evaluation of the variance factor can provide information on this matter. After computing the parameters to be estimated, their values can be substituted in the (non-linear) observation equations (2.44) to give the least squares corrections:

$$\hat{\mathbf{v}} = \mathbf{f}(\hat{\mathbf{x}}) - \mathbf{b} \tag{2.60}$$

the a posteriori variance factor $\hat{\sigma}_o^2$ can be estimated from

$$\hat{\sigma}_o^2 = (\hat{\mathbf{v}}^t\mathbf{W})/r \tag{2.61}$$

where r is the number of degrees of freedom in the LSE of $\hat{\mathbf{v}}$. For linearized observation equations, which is the usual case in close range photogrammetry, r is given by

$$r = m - u + d \tag{2.62}$$

where m is the number of observation equations, u is the number of parameters to be estimated and d is the number of minimum constraint equations needed to remove any defects in the co-ordinate datum. If the value of the a posteriori variance factor is equal to its a priori value of unity, then on average the standard deviations of the measurements conform to the design. Furthermore, there is no reason to doubt the overall adequacy of the chosen functional model. To decide whether or not $\hat{\sigma}_o^2 = 1$ the test

statistic $T = \hat{\mathbf{v}}^t\mathbf{W}\hat{\mathbf{v}}$ is used. It has a χ^2 distribution with r degrees of freedom if $\hat{\mathbf{v}}$is normally distributed with zero expectation. The test statistic T is used to test the null hypothesis H_0: $\hat{\sigma}_0^2 = 1$ against an alternative hypothesis. If $\hat{\sigma}_0^2 < 1$ the test is not normally carried out; it is accepted that the a priori variance factor indicates the design precision of the measurements has been exceeded and that the functional model is adequate. The alternative hypothesis is therefore not the two-tailed H_a: $\hat{\sigma}_0^2 \neq 1$, but the single-tailed H_a: $\hat{\sigma}_0^2 > 1$. The confidence level is usually 95% (significance level α = 5%). The critical value of the test statistic is $T_c = \chi^2_{r,\alpha}$. H_0 is accepted and H_a is rejected if $T < T_c$, or H_0 is rejected and H_a accepted if $T > T_c$. Rejection of H_0 means that any one or more of the following has occurred: the precision of the measurements has not met the design; there is a blunder associated with at least one measurement; and significant systematic error is present. The photogrammetrist must undertake further investigations and possibly re-measurement or re-calibration if possible. If and only if blunders and significant systematic errors are ruled out is it proper to use $\hat{\sigma}_0^2$ to scale cofactor matrices as in $\mathbf{C}_{\hat{x}} = \hat{\sigma}_0^2(\mathbf{A}^t\mathbf{W}\mathbf{A})^{-1}$. By doing this it is possible to show fully the effects on the variances of estimated parameters of a failure to meet the designed measurement precision. When the null hypothesis is accepted, cofactor matrices and covariance matrices are the same at significance level α (confidence level $1 - \alpha$).

An exception to the usual procedure occurs when, for the sake of simplicity in programming or data input the $m \times m$ identity matrix is used as the a priori weight matrix. This practice can be justified when the measurements are all of the same kind and made under the same conditions, which they sometimes are in purely photogrammetric networks with minimum constraints. Then the a posteriori variance factor (whose square root is in this case the standard deviation of an observation) must be used to scale cofactor matrices. If measurements are not all of the same type, or made under the same conditions, they may be separated into groups of the same type made under the same conditions and each group given the design standard deviation a priori. Tests of the a posteriori variance factor of each group of measurements are then made using *variance component estimation* (Caspary, 1988).

2.6.2 Indicators of precision

The covariance matrix $\mathbf{C}_{\hat{x}}$ contains all information necessary to define the precision of estimated target positions, camera exterior orientation elements, additional parameters and correlations between any pair of estimates, but it is not datum-invariant so the way in which the co-ordinate datum has been defined will influence the elements of $\mathbf{C}_{\hat{x}}$. If the total number of targets is t, the trace of the $3t \times 3t$ submatrix of $\mathbf{C}_{\hat{x}}$ ($= \mathbf{C}_{\hat{x}t}$ say) containing the variances and covariances of the estimated co-ordinates of all the targets can be used to give an indicator of the overall precision of position: $[(Tr\mathbf{C}_{\hat{x}t})/3t]^{1/2}$ is the square root of the mean variance of estimated target co-ordinates. Homogeneity of co-ordinate precision throughout the object space is often required in close range photogrammetry. The ratio $\lambda_{max}/\lambda_{min}$ of the largest eigenvalue of $\mathbf{C}_{\hat{x}t}$ to the smallest is a good indicator of this aspect of data quality; the closer the ratio is to unity, the more homogeneous is the precision of positioning. It must be recalled that $\mathbf{C}_{\hat{x}t}$ is singular in general with a rank deficiency of up to seven. If the zero variance reference base has been de-

fined by fixing co-ordinates, the corresponding variances, covariances and eigenvalues of $C_{\hat{x}\hat{x}}$ will be zero so generally they are not included in the two overall precision indicators given above.

Information about the precision of the estimated position of any target i is in the 3×3 sub-matrix of $C_{\hat{x}\hat{x}}$ (say $C_{\hat{x}_i\hat{x}_i}$) that contains the variances $\sigma_{X_i}^2$, $\sigma_{Y_i}^2$ and $\sigma_{Z_i}^2$ and covariances $\sigma_{X_iY_i}$, $\sigma_{Y_iZ_i}$ and $\sigma_{Z_iX_i}$ of the estimated co-ordinates of target i. The size and shape of the point standard triaxial ellipsoid for target i can be derived from the three eigenvalues and eigenvectors of $C_{\hat{x}_i\hat{x}_i}$; the lengths of three axes of the ellipsoid are the square roots of the three eigenvalues and their directions are those of the corresponding eigenvectors. The point standard triaxial ellipsoid for target i is a 20% *confidence region*. If the dimensions of the point standard ellipsoid are multiplied by a factor k, the level of confidence $(1-\alpha)$ of the corresponding three dimensional region is $P(\chi_3^2 < k^2)$. Alternatively, for each target i the size, shape and orientation of the point standard ellipse in each of the three co-ordinate planes can be computed algebraically from the elements of $C_{\hat{x}_i\hat{x}_i}$ (Cooper, 1987). If this is done for each of the t targets and the ellipses in a particular plane are displayed graphically, a good indication of the overall precision of target positions can be gained. Confidence levels corresponding to a scaling factor k of a point standard ellipse for two dimensional positioning are given by $P(\chi_2^2 < k^2)$.

The covariance matrix $C_{\hat{s}}$ includes the variances of estimated additional parameters, such as lens distortion parameters in a self calibrating bundle adjustment. The value of each estimated parameter should be tested to see whether or not it is different from zero. The null hypothesis is H_o: $\hat{p} = 0$ and the alternative hypothesis H_a: $\hat{p} \neq 0$ where $\hat{p}$ is the estimated value of the parameter being tested. The test statistic is $T = (p_o - \hat{p})/\sigma_{\hat{p}}$ where p_o is the true value of the parameter and $\sigma_{\hat{p}}$ is the standard deviation of $\hat{p}$. The test statistic T has a t distribution with f degrees of freedom, where f is the number of observation equations in which p appears, minus one. The test statistic is therefore $T = \hat{p}/\sigma_{\hat{p}}$. If the significance level of the test is α, H_o is accepted and H_a is rejected if $t_{f,1-\alpha/2} < T < t_{f,\alpha/2}$. Otherwise H_o is rejected and H_a is accepted. If the null hypothesis is accepted, the estimated parameter is zero at significance level α. It should be removed from the functional model and the LSE repeated; the parameter is unnecessary and its inclusion could complicate the detection of other defects in the functional model. 'Over-parametrization' can also occur when high correlation exists between pairs of estimated parameters. An effect could be to make the LSE process slow and in some cases impossible to obtain because the iterations fail to converge. The coefficient of correlation between any pair of parameters p and q is $\rho_{pq} = \sigma_{pq}/\sigma_p\sigma_q$ and its a priori value can be evaluated at the design stage so it should not be necessary to repeat the investigation after LSE. However, slow convergence of the iterative LSE process may indicate the presence of at least one pair of parameters with very high correlation, so an investigation is advisable. If a pair of parameters having a coefficient of correlation numerically greater than 0·9 is found, they are highly correlated (positively or negatively) and almost functionally related, so one of them should be removed from the functional model and the LSE process repeated.

2.6.3 *Tests for gross errors (Outliers)*

Each measurement b_i will have a least squares correction $\sigma_{\hat{v}_i}$ and a possible undetected

gross error Δb_i. The null hypothesis H_o: $\Delta b_i = 0$ can be tested against the alternative hypothesis H_a: $\Delta b_i \neq 0$ for each measurement in turn using the test statistic $T_i = \hat{v}_i / \sigma_o \sigma_{\hat{v}_i}$ which has a standard normal distribution $z \sim N(0,1)$ under H_o if the measurements **b** are independent and normally distributed with zero expectation (except for b_i) and if the variance factor is known a priori. If the significance level for the test is a, H_o is accepted and H_a is rejected if $z_{1-a/2} < T_i < z_{a/2}$. Otherwise H_o is rejected and H_a is accepted. Although none of the conditions for the validity of the test is likely to be fully satisfied in practice, most properly designed close range photogrammetric schemes meet them closely enough for the above test for the detection of gross errors to be useful. The variance of the least squares correction to the i^{th} measurement $\sigma_{\hat{v}_i}^2$ is the i^{th} diagonal element of $C_{\hat{v}}$ given by (2.58). The variance factor can be regarded as unity if the χ^2 test of the least squares corrections led to that conclusion. A χ^2 goodness-of-fit test may have shown that the least squares corrections are normally distributed with zero expectation. The τ distribution (Pope, 1976) can be used when the variance factor is not known a priori but is estimated as $\hat{\sigma}_o^2$. The test statistic is then $T_i = |\hat{v}_i| / \hat{\sigma}_o \sigma_{\hat{v}_i}$ which under H_o has the $\tau_{m,r}$ distribution where m is the number of measurements and r is the number of degrees of freedom in the estimation of $\hat{\sigma}_o^2$. The critical value for the test is $T_c = \tau_{m,r,a}$. If T_i is less than this value H_o is accepted and H_a is rejected. Tabulated values of $\tau_{m,r}$ are given by Cooper (1987).

The usual procedure is to evaluate each T_i and use the standard normal distribution to flag the measurements identified as having a gross error. These should then be examined in an attempt to identify a likely reason for the error, starting with the flagged measurement that has the highest value test statistic. If a good reason can be found, a measurement might be corrected or removed and the LSE process repeated to see if any improvements in data quality have taken place. Flagged measurements should not be discarded without additional evidence that they are in error. The errors might well be systematic and therefore capable of correction, either by calculation or by modification of the functional model and a new LSE.

An automated method for dealing with gross errors within the LSE process has come to be known as 'robustified LSE'. If, after convergence of the iterative LSE process, gross errors in the measurements are indicated by an unacceptably large value of the a posteriori variance factor and unacceptably large corrections to some measurements, the weights of those suspect measurements are decreased, usually by multiplying their a priori standard deviations by an explicit exponential function. Then the LSE process is repeated with the suspect measurements de-weighted. Tests of the a posteriori variance factor and of corrections to the measurements are again made, followed by further de-weighting and a new LSE process. The procedure is iterated until the a posteriori variance factor is insignificantly different from unity and all suspect measurements have been given such large standard deviations that they play little part in the LSE and are no longer flagged as having a gross error. They are effectively removed from the network. The method works well when a few gross errors are present, several times larger than the standard deviations of the corresponding measurements. In simulations it is often found (Caspary,1988 for example) that, at the end of the de-weighting procedure, the correction to a suspect measurement is almost equal to the (known) error in that measurement. However, in close range photogrammetry, error detection is most difficult

when systematic errors are suspected and automatic de-weighting does not then converge rapidly enough for it to be useful in that context.

2.6.4 Reliability

Reliability should be considered at the design stage to ensure adequate standards. Internal reliability can be assessed by computing a priori the marginally detectable gross error (MDGE) in each (proposed) measurement if the procedure described above is to be used. The MDGE is $\Delta b_i = F\sigma_{b_i}/r_i$ where F is a factor depending on the probabilities of Type I and Type II errors in the test for Δb_i, σ_{b_i} is the standard deviation of measurement i and r_i is the *redundancy number* of the measurement, i.e. the square root of the i^{th} diagonal element of $\mathbf{C}_v\mathbf{W}$. For example, if Δb_i is the only gross error present and the probabilities of Type I and Type II errors are 5% and 20% respectively, $F = 2{\cdot}8$ approximately. External reliability should also be examined at the design stage. The effect $\Delta\hat{\mathbf{x}}_i$ of an undetected MDGE Δb_i on the least squares estimate $\hat{\mathbf{x}}$ is

$$\Delta\hat{\mathbf{x}}_i = (\mathbf{A}^t\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^t\mathbf{W}\mathbf{e}_i\Delta b_i$$

where $\mathbf{e}_i$ is a $(m \times 1)$ vector with zero elements except for the i^{th} which is unity.

2.6.5 Data quality – Conclusion

There is no straightforward procedure to locate and remove what appear a posteriori to be errors. The photogrammetrist must consider instrumentation, procedures, conditions under which measurements were made, a priori statistics and algorithms as well as the functional model and its additional parameters when trying to identify errors. Detailed a priori design of each component of the measurement scheme, rigorous and regular calibration of equipment, as well as a posteriori analysis (automated where possible) must be carried out if data of acceptable precision, accuracy and reliability are to be made available quickly for further, non-photogrammetric processing. Photogrammetrists have the responsibility for devising methods and procedures to meet these requirements.

References and further reading

Abdel-Aziz, Y. and Karara, H.M., 1971. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. *Papers from the American Society of Photogrammetry Symposium on Close-Range Photogrammetry*, Urbana, Illinois.433 pages: 1-18.

Albertz, J. and Kreiling, W., 1975. *Photogrammetric guide.* Second Edition. Wichmann, Karlsruhe. 284 pages.

American Society for Photogrammetry and Remote Sensing, *passim*, monthly. *Photogrammetric Engineering and Remote Sensing.*

Bopp, H. and Krauss, H., 1978. An orientation and calibration method for non-topographic applications. *Photogrammetric Engineering and Remote Sensing*, 44(9):1191-1196.

Caspary, W.F., 1988. *Concepts of network and deformation analysis.* Monograph 11 (Ed. J.M. Rüeger) School of Surveying (now School of Geomatic Engineering), University of New South Wales, Australia. 183 pages.

Cooper, M.A.R., 1987. *Control surveys in civil engineering.* BSP Professional Books, Oxford (formerly Collins, London) and Nicholls, New York. 381 pages.

Cooper, M.A.R. and Cross, P.A., 1988. Statistical concepts and their application in photogrammetry and surveying. *Photogrammetric Record*, 12(71): 637–663.

Cooper, M.A.R. and Cross, P.A., 1991. Statistical concepts and their application in photogrammetry and surveying (continued). *Photogrammetric Record*, 13(77): 645–678.

Cox, A., 1966. *Photographic optics.* Focal Press, London. 490 pages.

International Society for Photogrammetry and Remote Sensing, *passim. International Archives of Photogrammetry and Remote Sensing.*

International Society for Photogrammetry and Remote Sensing, *passim* every two months. *ISPRS Journal of Photogrammetry and Remote Sensing.*

James, T.H. (Ed.), 1977. *The theory of the photographic process.* Fourth Edition. Macmillan, New York. 714 pages.

Karara, H.M., (Ed.) 1989. *Non-topographic photogrammetry.* Second Edition. American Society for Photogrammetry and Remote Sensing, Falls Church, Virginia. 445 pages.

Kennedy, J.B. and Neville, A.M., 1986. *Basic statistical methods for engineers and scientists.* Third Edition. Harper & Row, New York. 613 pages.

Kingslake, R., (Ed.) 1965-1969. *Applied optics and optical engineering*, Volumes 1-5, especially Vol 4:143-163. Academic Press, New York.

Koch, K.-R., 1988. *Parameter estimation and hypothesis testing in linear models.* Springer-Verlag, Berlin. 378 pages.

Kraus, K., 1993. *Photogrammetry. Volume 1. Fundamentals and standard processes.* Dümmler, Bonn. 397 pages.

Lawson, C.L. and Hanson, R.J., 1974. *Solving least squares problems.* Prentice-Hall, New Jersey. 340 pages.

Magill, A.A., 1955. Variation in distortion with magnification. *Journal of Research of the National Bureau of Standards*, 54(3):135–142.

Marzan, G.T. and Karara, H.M., 1975. A computer program for direct linear transformation solution of the colinearity condition, and some applications of it. *Proceedings of The American Society of Photogrammetry Symposium on Close-Range Photogrammetric Systems*, Champaign, Illinois. 670 pages: 420–476.

Mikhail, E.M., 1976. *Observations and least squares.* IEP, New York. 497 pages.

Photogrammetric Society, *passim*, bi-annually. *Photogrammetric Record.*

Pope, A.J., 1976. The statistics of residuals and the detection of outliers. *National Oceanic and Atmospheric Administration Technical Reports*, NOS65NGS1. US Department of Commerce, Washington. 133 pages.

Scott, P.J., 1977. The pupil in perspective. *Photogrammetric Record*, 9(49):83-92.

Shih, T.-Y., 1990. The duality and critical condition in the formulation and decomposition of a rotation matrix. *Photogrammetric Engineering and Remote Sensing*, 56(8):1173–1179.

Slama, C.C. (Ed.), 1980. *Manual of photogrammetry.* Fourth Edition. American Society of Photogrammetry, Falls Church, Virginia. 1056 pages.

Thompson, E.H., 1969. *Introduction to the algebra of matrices with some applications.* Adam Hilger, London. 229 pages.

Thompson, E.H., 1977. *Photogrammetry and surveying.* Photogrammetric Society, London. 359 pages.

Weatherburn, C.E., 1955. *Elementary vector analysis.* New and Revised Edition. G. Bell & Sons, London. 181 pages.

Zheng, Z. and Wang, X., 1992. A general solution of a closed-form space resection. *Photogrammetric Engineering and Remote Sensing*, 58(3):327–338.