Exploiting Multiple View Geometry in X-Ray Testing: Part I – Theory

Domingo Mery
Departamento de Ingeniería Informática
Universidad de Santiago de Chile
Av. Ecuador 3659, Santiago de Chile
dmery@ieee.org
http://www.diinf.usach.cl/∼dmery

September 12, 2003

Abstract

Multiple view geometry is increasingly being used in machine vision. In this paper we present a mathematical background of the multiple view geometry which is normally used in the X-ray computer vision. The article describes an explicit model which relates the 3D coordinates of an object to the 2D coordinates of the digital X-ray image pixel, the geometric and algebraic constraints between two, three and four X-ray images taken at different projections of the object, and the problem of 3D reconstruction from n views.

Keywords: X-Ray testing, automated inspection, computer vision, multiple view geometry.

1 Introduction

X-Ray testing is one of the more accepted ways for examining an object without destroying it. The purpose of this non-destructive method is to extract features of parts that are located inside the piece and are thus not detectable to the naked eye. A typical example is the inspection of castings [1].

The principle aspects of an automated X-ray inspection unit are shown in Fig. 1. Typically, it comprises the following five steps:

- The *manipulator* places the casting in the desired position.
- The *X-ray tube* generates X-rays which pass through the casting.
- The X-rays are detected by the fluorescent entrance screen of the *image intensifier*, amplified and depicted onto a phosphor screen. The image intensifier converts the X-rays to a visible radioscopic image.

- The guided and focussed image is registered by the *CCD-camera*.
- The *image processor* converts the analog video signal, transferred by the CCD-camera, into a digital data stream. Digital image processing is used to improve and evaluate the radioscopic image.

Currently, flat detectors made of amorphous silicon are being used as image sensors in some industrial inspection systems [2, 3]. In these detectors, the energy from the X-ray is converted directly into an electrical signal by a semi-conductor (without image intensifier). However, using flat detectors is not always feasible because of their high cost compared to image intensifiers.

In this article we present a background of the multiple view geometry which is normally used in the X-ray computer vision. We start presenting in Section 2, a model which relates the 3D coordinates of an object to the 2D coordinates of the digital X-ray image pixel. In Section 3, we establish the geometric and algebraic constraints between two, three and four X-ray images obtained as different projections of the object. The problem of the 3D reconstruction is explained in Section 4. Applications of the multiple view X-ray testing can be found in [4].

2 Geometric model

In this Section we present a model which relates the 3D coordinates of an object to the 2D coordinates of the image pixel. The object will be examined from n different viewpoints. In this approach, *homogeneous coordinates* [5] are used: a point $(a_1, a_2, ..., a_N)$ in a N dimensional space is expressed as a homogeneous vector with $N+1$ elements $(b_1, b_2, ..., b_N, b_{N+1})$ where $a_i = b_i/b_{N+1}$ for $i = 1, ...N$. Table 1 shows the transformations in 2D and 3D.

2.1 Coordinate systems

Since the X-ray images are taken at n different positions, we use an index p, $p = 1, ..., n$, to denote the p-th position.

Let us define the following coordinate systems to describe the relationship between 3D object point and 2D pixel, as shown in Figure 1:

• **Object coordinate system:** The 3D object coordinate system is attached to the object. An object point M in this coordinate system is denoted by

$$\mathbf{M} = [X \ Y \ Z \ 1]^\mathsf{T} \tag{1}$$

in homogeneous coordinates. We use the notation of Faugeras [5], where we differentiate between the projective geometric objects themselves and their representations, e.g. a point in the space will be denoted by M whereas its vector in homogeneous coordinates will be denoted by $\mathbf{M}$. The centre of rotation of the object is assumed to be at the origin O of this coordinate system. The motion

of the object is considered as a rotation around the origin, followed by a translation. The coordinates of M in the object coordinate system is independent of the object displacement, i.e. $\mathbf{M} = \mathbf{M}_p$.

• **World coordinate system:** The 3D world coordinate system is defined in the optical centre of the central projection, i.e. its origin C corresponds to the X-ray source modelled as a point. The object point M at its p-th position in this coordinate system is

$$\bar{\mathbf{M}}_p = [\bar{X}_p \ \bar{Y}_p \ \bar{Z}_p \ 1]^\mathsf{T} \tag{2}$$

in homogeneous coordinates. The object coordinate system is then considered as a rigid displacement of the world coordinate system represented by a 3×1 translation vector $\mathbf{t}_p = [t_X \ t_Y \ t_Z]_p^\mathsf{T}$ and a 3×3 rotation matrix:

$$\mathbf{R}(\omega_X, \omega_Y, \omega_Z)_p = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}_p , \tag{3}$$

where the elements R_{ij} can be expressed as a function of cosinus and sinus of the Euler angles ω_X, ω_Y, and ω_Z that describe the rotation of the X, Y and Z axes respectively [6]:

$$\begin{aligned}
R_{11} &= \cos(\omega_Y)\cos(\omega_Z) \\
R_{12} &= \cos(\omega_Y)\sin(\omega_Z) \\
R_{13} &= -\sin(\omega_Y) \\
R_{21} &= \sin(\omega_X)\sin(\omega_Y)\cos(\omega_Z) - \cos(\omega_X)\sin(\omega_Z) \\
R_{22} &= \sin(\omega_X)\sin(\omega_Y)\sin(\omega_Z) + \cos(\omega_X)\cos(\omega_Z) \ . \\
R_{23} &= \sin(\omega_X)\cos(\omega_Y) \\
R_{31} &= \cos(\omega_X)\sin(\omega_Y)\cos(\omega_Z) + \sin(\omega_X)\sin(\omega_Z) \\
R_{32} &= \cos(\omega_X)\sin(\omega_Y)\sin(\omega_Z) - \sin(\omega_X)\cos(\omega_Z) \\
R_{33} &= \cos(\omega_X)\cos(\omega_Y)
\end{aligned}$$

Finally, with the 4×4 matrix $\mathbf{D}_p$ and its inverse:

$$\mathbf{D}_p = \begin{bmatrix} \mathbf{R}_p & \mathbf{t}_p \\ \mathbf{0} & 1 \end{bmatrix} \qquad \text{and} \qquad \mathbf{D}_p^{-1} = \begin{bmatrix} \mathbf{R}_p^\mathsf{T} & -\mathbf{R}_p^\mathsf{T}\mathbf{t}_p \\ \mathbf{0} & 1 \end{bmatrix}, \tag{4}$$

we obtain the relationships between object and world coordinate system:

$$\bar{\mathbf{M}}_p = \mathbf{D}_p\mathbf{M} \qquad \text{and} \qquad \mathbf{M} = \mathbf{D}_p^{-1}\bar{\mathbf{M}}_p. \tag{5}$$

• **X-ray projection coordinate system:** We define a 2D projection coordinate system that indicates the coordinates of a point in the (no visible) X-ray image at a fictitious plane $\bar{Z} = f$ located at the entrance screen of the image intensifier, where f is equivalent to the optical focal length. Its origin o is pierced by the optical axis ($\bar{Z}$-axis). The p-th projection of the object point M in this coordinate system, denoted by m_p, in homogeneous coordinates is

$$\mathbf{m}_p = [x_p \ y_p \ 1]^\mathsf{T}. \tag{6}$$

The X-rays perform a linear perspective projection of the point M onto a point m_p in the fictitious plane without any distortion. Using $\bar{\mathbf{M}}_p$ and $\mathbf{m}_p$ as homogeneous representations of M and m_p respectively, we have the linear equation:

$$\lambda_p \mathbf{m}_p = \underbrace{\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{B}} \bar{\mathbf{M}}_p \tag{7}$$

where λ_p is a scale factor. Let us denote by $\mathbf{P}_p = \mathbf{B}\mathbf{D}_p$ the p-th 3×4 perspective projection matrix. From (5) and (7) we obtain the equation that maps object coordinates to projection plane coordinates at the the p-th position:

$$\lambda_p \mathbf{m}_p = \mathbf{P}_p \mathbf{M}. \tag{8}$$

• **Digital image coordinate system:** Now, we introduce the 2D image coordinate system as a representation of the pixel coordinates of the (visible) X-ray image formed at the CCD camera or at the flat panel detector. A point m_p is projected onto the plane of the digital image array as w_p which representation in this coordinate system is

$$\mathbf{w}_p = [u_p \; v_p \; 1]^\mathsf{T} \tag{9}$$

in homogeneous coordinates. Using flat detectors, the transformation between $\mathbf{m}_p$ and $\mathbf{w}_p$ can be modelled according to:

$$\mathbf{w}_p = \mathbf{T}\mathbf{m}_p \qquad \text{and} \qquad \mathbf{m}_p = \mathbf{T}^{-1}\mathbf{w}_p, \tag{10}$$

where matrix $\mathbf{T}$ is a homogeneous 3×3 matrix that causes a general 2D perspective transformation where rotation, translation, scaling, skew and perspective distortion are considered [7].

Nevertheless, the transformation $\mathbf{m}_p \rightarrow \mathbf{w}_p$ is non-linear in radioscopic systems. Due to the curvature of the entrance screen of the image intensifier, the radioscopic image is deformed, specially at the corners of the image. Therefore, the relationship between projection and image coordinate system can be expressed by:

$$\mathbf{w}_p = \mathbf{f}(\mathbf{m}_p) \qquad \text{and} \qquad \mathbf{m}_p = \mathbf{f}^{-1}(\mathbf{w}_p). \tag{11}$$

The non-linear function $\mathbf{f}$ can be modelled as hyperbolic [8, 9, 10], cubic [11] or as a radial and decentring distortion [12]. A way to estimate this function is analysing the projective distortion of a calibration plate which contains holes placed in a regular grid manner. The phenomenon of the distortion effect is illustrated in the radioscopic image of the calibration plate, as shown in Fig. 2a.

The hyperbolic model used to transform the 2D point $\mathbf{m}_p = [x_p \; y_p \; 1]^\mathsf{T}$ (in the X-ray projection coordinate system) into the 2D point $\mathbf{w}_p = [u_p \; v_p \; 1]^\mathsf{T}$ (in the digital image coordinate system) will be explained next. The input screen

of the image intensifier corresponds normally to a hyperbolic 3D surface [13], which can be defined by:

$$\bar{Z} = F(\bar{X}, \bar{Y}) = f\sqrt{1 + (\bar{X}/a)^2 + (\bar{Y}/b)^2}, \tag{12}$$

with f being the real half axis of the hyperboloid; and a and b the imaginary half axes. We observe that f coincides with the focal length of the X-ray projection. As shown in Fig. 3, the projection of point M onto the input screen of the image intensifier is denoted by m'_p. It is calculated as the intersection of the line that contains points C, M and m_p with the 3D surface F. Its coordinates are given by:

$$\mathbf{m'}_p = \mathbf{g}(\mathbf{m}_p) = [cx_p \; cy_p \; 1]^\mathsf{T}, \tag{13}$$

with $c = 1/\sqrt{1 - (x_p/a)^2 - (y_p/b)^2}$. The point m'_p is imaged at the CCD camera as w_p, the coordinates of which can be estimated approximately using an affin transformation [5]:

$$\mathbf{w}_p = \underbrace{\begin{bmatrix} +k_x\cos(\theta) & +k_y\sin(\theta) & u_0 \\ -k_x\sin(\theta) & +k_y\cos(\theta) & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \mathbf{m'}_p \tag{14}$$

where k_x and k_y are scale factors, and (u_0, v_0) and θ are respectively the translation and the rotation between x, y– and u, v–axes. The transformation from $\mathbf{m}_p$ to $\mathbf{w}_p$ is computed by:

$$\mathbf{w}_p = \mathbf{f}(\mathbf{m}_p) = \mathbf{A}\mathbf{g}(\mathbf{m}_p), \tag{15}$$

and the inverse transformation of $\mathbf{f}$ is given by:

$$\mathbf{m}_p = \mathbf{f}^{-1}(\mathbf{w}_p) = \mathbf{g}^{-1}(\mathbf{A}^{-1}\mathbf{w}_p) = [dx'_p \; dy'_p \; 1]^\mathsf{T} \tag{16}$$

with $d = 1/\sqrt{1 + (x'_p/a)^2 + (y'_p/b)^2}$. The modelled hyperbolic grid is shown in Fig. 2b.

To summarise, using (8) for the perspective projection and (15) –for the projection in image intensifiers– or (10) –for for the projection in flat detectors– , an object point M, whose homogeneous coordinates are $\mathbf{M} = [X \; Y \; Z \; 1]^\mathsf{T}$ in a coordinate system attached to the object, can be mapped into a 2D point of the digital X-ray image as w_p (in position p), the coordinates of which are:

$$\mathbf{w}_p = [u_p \; v_p \; 1]^\mathsf{T} = \mathbf{F}(\mathbf{\Theta}_p, \mathbf{M}), \tag{17}$$

where $\mathbf{\Theta}_p$ is the vector of parameters involved in the projection model.

2.2 Calibration

The calibration of an X-ray imaging system –in the context of 3D machine vision– is the process of estimating the parameters of a model, which is used to

determine the projection of the 3D object under test into its 2D digital X-ray image. This relationship 3D $\rightarrow$ 2D can be modelled with the transfer function $\mathbf{F} : \mathcal{R}^3 \rightarrow \mathcal{R}^2$ expressed in (17).

There are several calibration techniques developed by the computer vision community to calibrate an imaging system. They can be roughly classified into two categories: *photogrammetric calibration* and *self–calibration* [14]. The first one is a 3D reference object–based calibration, where the calibration is performed by observing a calibration object whose geometry in 3D space is known [5]. The second technique uses the identification of matching points in several views of a scene taken by the same camera. Self–calibration does not use a calibration object with known 3D geometry because it aims to identify the intrinsic parameters of the camera and to reconstruct 3D structure up to a scale similarity [15, 16].

Due to the high precision feature measurement of 3D geometry required in the NDE applications, it would be necessary to do a *true* reconstruction of the 3D space without a scale factor. For this reason, the calibration technique used in NDE belongs to the photogrammetric category: it uses a 3D reference object with known geometry.

The parameters of the model are estimated by minimising an objective function, which is the distance between the modelled projection of a set of object points with its measured projection. Thus, we estimate the parameters of the model based on a number of points whose object coordinates (X, Y, Z) are known and whose image coordinates (u, v) are measured. We define the m calibration points of the calibration object as $\mathbf{M}_i = [X_i \ Y_i \ Z_i \ 1]^\mathsf{T}$, for $i = 1, ..., m$. Using (17) we obtain the *reprojected* points w_{ip}, that are the inferred projections in image p computed from the calibration points using the calibration parameters $\mathbf{\Theta}_p$, for $p = 1, ..., n$. The parameter vector is then computed by minimising the distance between measured points ($\mathbf{w}_{ip}$) and estimated points ($\hat{\mathbf{w}}_{ip} = \mathbf{F}(\mathbf{\Theta}_p, \mathbf{M}_i)$). The objective function $J(\mathbf{\Theta})$ is defined as the mean-square discrepancy between these points:

$$J(\mathbf{\Theta}) = \frac{1}{n\ m} \sum_{p=1}^{n} \sum_{i=1}^{m} \parallel \mathbf{w}_i - \mathbf{F}(\mathbf{\Theta}_p, \mathbf{M}_i) \parallel \rightarrow \min, \tag{18}$$

where $\mathbf{\Theta} = [\mathbf{\Theta}_1 ... \mathbf{\Theta}_n]^\mathsf{T}$. The calibration problem is a non-linear optimisation problem, where the minimisation of $J(\mathbf{\Theta})$ has no closed-form solution[1]. For this reason, the objective function must be iteratively minimised starting with an initial estimated value $\mathbf{\Theta}^0$.

3 Corresponding points in multiple view

Corresponding points are those projection points (in different views) that represent projections of the same 3D point. With the mathematical model of Section

[1] Under some assumptions a closed-form solution can be obtained if the geometric distortion is not considered [17, 12].

2.1, we can relate the 3D coordinates of the object into the digital X-ray image pixel coordinates at each projection of the object. In this Section, we consider geometric and algebraic constraints to solve the correspondence problem between X-ray images obtained as different projections of the test object. In the approach, a corresponding point w_p, that is found in the p-th image, is firstly transformed into the point m_p of the X-ray projection coordinate system using the linear transformation (10) for flat detectors or the non-linear approach (11) or (16) for image intensifiers. With this transformation we can use the linear relationship (8) explained above.

3.1 Correspondence between two views

Now, the correspondence between two points m_p and m_q (in the X-ray projection coordinate system) is considered. The first point is obtained by projecting the object point M at position p, and the second one at position q.

To solve the correspondence problem we use epipolar geometry [5, 8]. The epipolar constraint is well known in stereo vision: for each projection point m_p at the position p, its corresponding projection point m_q at the position q lies on the epipolar line ℓ of m_p, as shown in Fig. 4, where C_p and C_q are the centers of projections p and q respectively. In this representation, a rotation and translation relative to the object coordinate system is assumed. The epipolar line ℓ can be calculated as the projection of line $\langle m_p, C_p \rangle$ by the center of projection C_q into projection plane q.

Usually, the epipolar line can be obtained as follows: m_p determines a ray between itself and optical centre C_p (X-ray source). The coordinates of these points in the world coordinate system are:

$$\bar{\mathbf{C}}_\mathbf{p} = [0\ 0\ 0\ 1]^\mathsf{T} \qquad \text{and} \qquad \dot{\bar{\mathbf{m}}}_p = [x_p\ y_p\ f\ 1]^\mathsf{T}$$

respectively. The bar denotes that the coordinates are given in the world coordinate system, and the dot over $\mathbf{m}$ indicates that the projection point m is given in a 3D coordinate system instead of a 2D coordinate system. $\dot{\bar{\mathbf{m}}}_p$ can be expressed by $\dot{\bar{\mathbf{m}}}_p = \mathbf{G}\mathbf{m}_p$ where

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & f \\ 0 & 0 & 1 \end{bmatrix}.$$

Using (4), the vectors $\bar{\mathbf{C}}_\mathbf{p}$ and $\dot{\bar{\mathbf{m}}}_p$ can be transformed into coordinates of the object coordinate system as:

$$\mathbf{C}_p = \mathbf{D}_p^{-1}\bar{\mathbf{C}}_\mathbf{p} = [-\mathbf{t}_p^\mathsf{T}\mathbf{R}_p\ 1]^\mathsf{T} \qquad \text{and} \qquad \dot{\mathbf{m}}_p = \mathbf{D}_p^{-1}\mathbf{G}\mathbf{m}_p.$$

These points can be projected into plane q using (8):

$$\mathbf{e}_{pq} = \mathbf{P}_q[-\mathbf{t}_p^\mathsf{T}\mathbf{R}_p\ 1]^\mathsf{T} = [e_x\ e_y\ e_z]^\mathsf{T} \qquad \text{and} \qquad \mathbf{m}_{pq} = \mathbf{P}_q\mathbf{D}_p^{-1}\mathbf{G}\mathbf{m}_p.$$

The first one, i.e. the projection of C_p into plane q, is the well known *epipole*[2]. The epipolar line is defined as the line that contains the epipole $\mathbf{e}_{pq}$ and the point $\mathbf{m}_{pq}$. We observe that, since the epipole depends only on the two views geometry (and not on the points m_p and m_q), the epipole belongs to any epipolar line obtained from an arbitrary point m_p.

The projective representation of the epipolar line is obtained by taking the cross-product of these two points, i.e. $\ell = \mathbf{e}_{pq} \times \mathbf{m}_{pq}$. Line ℓ can be written using $[\mathbf{e}_{pq}]_\times$, the anti-symmetric matrix of $\mathbf{e}_{pq}$, where $\ell = [\mathbf{e}_{pq}]_\times \mathbf{m}_{pq}$. $[\mathbf{e}_{pq}]_\times$ is defined as the 3×3 matrix such that $[\mathbf{e}_{pq}]_\times \mathbf{s} = \mathbf{e}_{pq} \times \mathbf{s}$ for all vectors $\mathbf{a}$, i.e.

$$[\mathbf{e}_{pq}]_\times = \begin{bmatrix} 0 & e_z & -e_y \\ -e_z & 0 & e_x \\ e_y & -e_x & 0 \end{bmatrix}.$$

Thus, line ℓ is computed by

$$\ell = \mathbf{E}_{pq}\mathbf{m}_p \tag{19}$$

where $\mathbf{E}_{pq}$ is the well known *Essential Matrix* [18] given by:

$$\mathbf{E}_{pq} = [\mathbf{e}_{pq}]_\times \mathbf{P}_q \mathbf{D}_p^{-1} \mathbf{G}. \tag{20}$$

Since the point m_q belongs to the epipolar line ℓ, it follows that

$$\mathbf{m}_q^\mathsf{T} \ell = \mathbf{m}_q^\mathsf{T} \mathbf{E}_{pq} \mathbf{m}_p = \ell_x x_q + \ell_y y_q + \ell_z = 0 \tag{21}$$

with $[\ell_x \ \ell_y \ \ell_z]^\mathsf{T} = \mathbf{E}_{pq}\mathbf{m}_p$. Equation (21) is known as the *epipolar constraint*: If m_p and m_q are corresponding points, then m_q must lie on the epipolar line ℓ of m_p, i.e. $\mathbf{m}_q^\mathsf{T}\mathbf{E}_{pq}\mathbf{m}_p$ must be zero.

In practice, the projection points m_p and m_q can be corresponding points, if the perpendicular Euclidean distance from the epipolar line ℓ of the point m_p to the point m_q is smaller than a small number ε_2 [19]:

$$d_2 = \frac{|\mathbf{m}_q^\mathsf{T} \mathbf{E}_{pq} \mathbf{m}_p|}{\sqrt{\ell_1^2 + \ell_2^2}} < \varepsilon_2. \tag{22}$$

Another way to estimate the epipolar constraint is using *bifocal tensors* [20, 21], as explained next. From (8) the two projections can be expressed a by:

$$\begin{cases} \lambda_p \mathbf{m}_p &= \mathbf{P}_p \mathbf{M} &:= \mathbf{A}\mathbf{M} \\ \lambda_q \mathbf{m}_q &= \mathbf{P}_q \mathbf{M} &:= \mathbf{B}\mathbf{M} \end{cases}. \tag{23}$$

These two equations can also be written as:

$$\underbrace{\begin{bmatrix} \mathbf{a}_1 & x_p & 0 \\ \mathbf{a}_2 & y_p & 0 \\ \mathbf{a}_3 & 1 & 0 \\ \mathbf{b}_1 & 0 & x_q \\ \mathbf{b}_2 & 0 & y_q \\ \mathbf{b}_3 & 0 & 1 \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} \mathbf{M} \\ -\lambda_p \\ -\lambda_q \end{bmatrix}}_{\mathbf{v}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{24}$$

[2]The word *epipole* comes from the Greek ἐπι (*epi*): over and πόλος (*polos*): attractor.

where $\mathbf{a}_i$ and $\mathbf{b}_i$ denote the i-th row of matrices $\mathbf{A}$ and $\mathbf{B}$ respectively. If m_p and m_q are corresponding points, then the 3D point M exists. It follows that there must be a nontrivial solution for $\mathbf{v}$ in (24), i.e. the determinant of the 6×6 matrix $\mathbf{G}$ must be zero. Expanding the determinant of $\mathbf{G}$ we obtain:

$$|\mathbf{G}| = [x_q \; y_q \; 1] \underbrace{\begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix}}_{\mathbf{F}_{pq}} \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \mathbf{m}_q^{\mathsf{T}} \mathbf{F}_{pq} \mathbf{m}_p = 0 \qquad (25)$$

where $\mathbf{F}_{pq}$ corresponds to the mentioned Essential-Matrix of equation (20) when $\mathbf{A} = \mathbf{P}_p$ and $\mathbf{B} = \mathbf{P}_q$. Generally, the projection matrices $\mathbf{A}$ and $\mathbf{B}$ can be arbitrary 3×4 projection matrices. In this case, the Essential-Matrix is called the Fundamental Matrix [7, 22], and its elements, called *bifocal tensors*, can be computed as:

$$F_{ij} = (-1)^{i+j} \begin{vmatrix} \sim \mathbf{a}_j \\ \sim \mathbf{b}_i \end{vmatrix} \qquad \text{for } i, j = 1, 2, 3. \qquad (26)$$

where $\sim \mathbf{a}_j$ and $\sim \mathbf{b}_i$ mean respectively matrix $\mathbf{A}$ without the j-th row and matrix $\mathbf{B}$ without the i-th row.

Usually, we can express matrix $\mathbf{A}$ in a canonical form:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = [\mathbf{I} \mid \mathbf{0}]. \qquad (27)$$

The canonical form can be achieved using a general projective transformation of the object coordinate system: $\mathbf{M}' = \mathbf{H}\mathbf{M}$, where $\mathbf{M}'$ is the transformation of $\mathbf{M}$, and $\mathbf{H}$ is a 4×4 non-singular matrix obtained by adding one extra row to $\mathbf{P}_p$ [23]. Thus, equation (23) can be expressed as:

$$\begin{cases} \lambda_p \mathbf{m}_p = [\mathbf{I} \mid \mathbf{0}]\mathbf{M}' = \mathbf{A}\mathbf{M}' \\ \lambda_q \mathbf{m}_q = \mathbf{B}\mathbf{M}' \end{cases} \qquad (28)$$

with

$$\begin{array}{rcl} \mathbf{M}' & = & \mathbf{H}\mathbf{M} \\ \mathbf{A} & = & \mathbf{P}_p\mathbf{H}^{-1} \\ \mathbf{B} & = & \mathbf{P}_q\mathbf{H}^{-1} \end{array} .$$

For the canonical form $\mathbf{A} = [\mathbf{I} \mid \mathbf{0}]$, the bifocal tensors may be expressed by:

$$F_{ij} = b_{i\oplus 1,j} b_{i\oplus 2,4} - b_{i\oplus 2,j} b_{i\oplus 1,4} \qquad (29)$$

where

$$i \oplus k = \begin{cases} i + k & \text{if } i + k \leq 3 \\ i + k - 3 & \text{otherwise} \end{cases} .$$

An example of corresponding points is illustrated in Fig. 5. In the left view points are marked. Using the non-linear model (11) and the epipolar constraint (25) the epipolar curves given by

$$[\mathbf{f}^{-1}(\mathbf{w}_q)]^{\mathsf{T}}\mathbf{F}_{pq}[\mathbf{f}^{-1}(\mathbf{w}_p)] = 0 \tag{30}$$

are shown in the right view. We observe that the corresponding points in second view lie on the epipolar curves.

An additional criterion to establish the correspondence between two views is that the 3D point reconstructed from the projection points m_p and m_q must belong to the space occupied by the test object [10]. From m_p and m_q the corresponding 3D point $\hat{M}$ can be estimated using 3D reconstruction techniques (see Section 4). It is necessary to examine if $\hat{M}$ resides in the volume of the test object, the dimensions of which are usually known a priori (e.g. a wheel is assumed to be a cylinder). This criterion implies that the epipolar is delimited as illustrated in Fig. 6. It is possible to use a CAD model of the test object to evaluate this criterion in a more precise way.

3.2 Correspondence between three views

In the three views case we have the projection points m_p, m_q and m_r at p-th, q-th and r-th positions respectively. The correspondence in three views can be established by calculating the epipolar lines of m_p and m_q in third view as shown in Fig. 7. If the intersection coincides with m_r, then the three points are corresponding. However, the intersection of epipolar lines in trifocal geometry is not well-defined when the epipolar lines are equal. This situation occurs in two cases: *i)* when the 3D point M that has generated the points m_p, m_q and m_r, lie in the plane defined by the three optical centres, and *ii)* when the three optical centres are aligned [24].

In order to avoid these singularities, the relationships in three views are generally described using *trifocal tensors* [7]. Analogous to the two views case explained in Section 3.1, the three projection equations are:

$$\left\{\begin{array}{lcll} \lambda_p\mathbf{m}_p & = & \mathbf{P}_p\mathbf{M} & := \quad \mathbf{AM} \\ \lambda_q\mathbf{m}_q & = & \mathbf{P}_q\mathbf{M} & := \quad \mathbf{BM} \\ \lambda_r\mathbf{m}_r & = & \mathbf{P}_r\mathbf{M} & := \quad \mathbf{CM} \end{array}\right. . \tag{31}$$

They can be written according to (24) by:

$$\underbrace{\begin{bmatrix} \mathbf{a}_1 & x_p & 0 & 0 \\ \mathbf{a}_2 & y_p & 0 & 0 \\ \mathbf{a}_3 & 1 & 0 & 0 \\ \mathbf{b}_1 & 0 & x_q & 0 \\ \mathbf{b}_2 & 0 & y_q & 0 \\ \mathbf{b}_3 & 0 & 1 & 0 \\ \mathbf{c}_1 & 0 & 0 & x_r \\ \mathbf{c}_2 & 0 & 0 & y_r \\ \mathbf{c}_3 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} \mathbf{M} \\ -\lambda_p \\ -\lambda_q \\ -\lambda_r \end{bmatrix}}_{\mathbf{v}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{32}$$

where $\mathbf{c}_i$ denotes the i-th row of matrix $\mathbf{C} = \mathbf{P}_r$. If m_p, m_q and m_r are corresponding points, then there must be a nontrivial solution for $\mathbf{v}$. It follows that the rank of the 9×7 matrix $\mathbf{G}$ must be at most 6. In other words, all 7×7 submatrices have vanishing determinants. The minors of $\mathbf{G}$ can be written using Laplace expansions as sums of products of determinants of four rows taken form the first four columns of $\mathbf{G}$ and products of image coordinates [25]. By expanding the determinants, we can find four linearly independent relationships:

$$\left\{ \begin{array}{lrcl} D_1 = & (x_r\mathbf{T}^{13} - x_r x_q \mathbf{T}^{33} + x_q \mathbf{T}^{31} - \mathbf{T}^{11})\mathbf{m}_p & = & 0 \\ D_2 = & (y_r\mathbf{T}^{13} - y_r x_q \mathbf{T}^{33} + x_q \mathbf{T}^{32} - \mathbf{T}^{12})\mathbf{m}_p & = & 0 \\ D_3 = & (x_r\mathbf{T}^{23} - x_r y_q \mathbf{T}^{33} + y_q \mathbf{T}^{31} - \mathbf{T}^{21})\mathbf{m}_p & = & 0 \\ D_4 = & (y_r\mathbf{T}^{23} - y_r y_q \mathbf{T}^{33} + y_q \mathbf{T}^{32} - \mathbf{T}^{22})\mathbf{m}_p & = & 0 \end{array} \right., \tag{33}$$

where

$$\mathbf{T}^{jk} = [T_1^{jk}\ T_2^{jk}\ T_3^{jk}],$$

and

$$T_i^{jk} = (-1)^{i+1} \begin{vmatrix} \sim \mathbf{a}_i \\ \mathbf{b}_j \\ \mathbf{c}_k \end{vmatrix} = \qquad \text{for } i,j,k = 1,2,3, \tag{34}$$

where $\sim \mathbf{a}_i$ means the matrix $\mathbf{A}$ without row i. The elements T_i^{jk} are called the *trifocal tensors* for the images p, q and r [23, 20]. For the canonical form $\mathbf{A} = [\mathbf{I} \mid \mathbf{0}]$, the trifocal tensors may be easily obtained by:

$$T_i^{jk} = b_{ji}c_{k4} - b_{j4}c_{ki} \qquad \text{for } i,j,k = 1,2,3. \tag{35}$$

The equations denoted by (33) above, are known as the *Trilinearities of Shashua* [26]. They establish a linear relationship between the coordinates of points m_p, m_q y m_r to find the correspondence. If the three points satisfy the four trilinearities, then they are corresponding points. Equation (34) implies that the trifocal tensors do not depend on the points of the images, they are computed from the three projection matrices.

The *reprojection* of m_r, i.e. the coordinates $\hat{x}_r$ and $\hat{y}_r$ obtained from the points m_p and m_q, may be simply estimated from the trilinearities (33):

$$\alpha\hat{\mathbf{m}}_r = (\mathbf{T}^1 - x_q\mathbf{T}^3)\mathbf{m}_p = (\mathbf{T}^2 - y_q\mathbf{T}^3)\mathbf{m}_p. \tag{36}$$

where α is a scale factor, $\hat{\mathbf{m}}_r = [\hat{x}_r\ \hat{y}_r\ 1]^\mathsf{T}$, and $\mathbf{T}^j$ is a 3×3 matrix with the (k, i)-element equal to T_i^{jk}.

In practice, given two corresponding points $\mathbf{m}_p$ and $\mathbf{m}_q$, the third one $\mathbf{m}_r$ can be considered as the corresponding point in third view, if the Euclidean distance between $\mathbf{m}_r$ and its reprojection $\hat{\mathbf{m}}_r$ is smaller than a small number ε_3:

$$d_3 = \|\hat{\mathbf{m}}_r - \mathbf{m}_r\| < \varepsilon_3. \tag{37}$$

3.3 Correspondence between four views

In the four views case we have the projection points m_p, m_q, m_r and m_s at p-th, q-th r-th and s-th positions respectively. Similar to the previous sections we can write the four projection equations as a linear equation $\mathbf{Gv} = \mathbf{0}$. Once more, the existence of a nontrivial solution for $\mathbf{v}$ yields in this case to the condition that all 8×8 minors of $\mathbf{G}$ must be zero. Thus, we obtain the well known 81 *quadrifocal tensors* and the corresponding 16 *quadrilinearities* [25, 20].

In practice, the quadrilinearities are not used because they are redundant. Corresponding constraints in four views are obtained from the trilinearities. Thus, the points m_p, m_q, m_r and m_s are corresponding if m_p, m_q and m_r are corresponding, and m_q, m_r and m_s are corresponding as well [27].

4 Three-dimensional reconstruction

In NDT & E, three-dimensional reconstruction is usually related to computed tomography (CT). However, in Computer Vision the attempt is made to estimate only the location (and not the X-ray absorption coefficient) of 3D points in space. In this sense, the reconstruction is based on photogrammetric rather than tomographic methods.

In this Section, two approaches that perform the 3D reconstruction, in sense of locating in 3D space, will be described. The 3D reconstruction will be done from corresponding points in X-ray projection coordinate system. As explained in Section 3, a point w_p, that is found in the p-th image, is first transformed into the coordinates m_p of the X-ray projection coordinate system.

4.1 Linear 3D reconstruction from two views

We estimate now the 3D point M from two corresponding points m_p and m_q using the linear approach introduced by Hartley [23]. Without loss of generality, the method employs the canonical form (see equation (28)) for the first projection:

$$\lambda_p\mathbf{m}_p = [\mathbf{I} \mid \mathbf{0}]\mathbf{M}'.$$

Thus, the transformed 3D point can be expressed by:

$$\mathbf{M}' = \lambda_p[\mathbf{m}_p^\mathsf{T}\ 1/\lambda_p]^\mathsf{T}. \tag{38}$$

The second projection of this point yields

$$\lambda_q\mathbf{m}_q = \mathbf{B}\mathbf{M}' = \mathbf{B}\lambda_p[\mathbf{m}_p^\mathsf{T}\ 1/\lambda_p]^\mathsf{T}, \tag{39}$$

that is an equation system with three linear equations in the unknowns λ_p and λ_q. If m_p and m_q are corresponding points one may consider only two of these three equations. Taking for example the first two equations one may compute λ_p. Substituting the value of λ_p into (38) and after some simplifications we obtain:

$$\mathbf{M} = \mathbf{H}^{-1}\mathbf{M}' = \alpha\mathbf{H}^{-1}\begin{bmatrix} (y_q b_{14} - x_q b_{24})\mathbf{m}_p \\ (x_q\mathbf{b}_2 - y_q\mathbf{b}_1)\mathbf{m}_p \end{bmatrix} \tag{40}$$

where α is a scale factor.

4.2 3D reconstruction from two or more views

We assume that we have $n > 1$ projections at n different positions. In these projections we have found the corresponding points m_p, $p = 1, ..., n$, with coordinates (x_p, y_p). To reconstruct the corresponding 3D point M that has produced these projection points, we use a least squares technique [5].

Each projection yields the equation (8), with three linear equations in the unknowns (X, Y, Z) and λ_p:

$$\begin{bmatrix} \lambda_1 x_1 \\ \lambda_1 y_1 \\ \lambda_1 \\ \vdots \\ \lambda_n x_n \\ \lambda_n y_n \\ \lambda_n \end{bmatrix} = \begin{bmatrix} s_{11}^1 & s_{12}^1 & s_{13}^1 & s_{14}^1 \\ s_{21}^1 & s_{22}^1 & s_{23}^1 & s_{24}^1 \\ s_{31}^1 & s_{32}^1 & s_{33}^1 & s_{34}^1 \\ \vdots & \vdots & \vdots & \vdots \\ s_{11}^n & s_{12}^n & s_{13}^n & s_{14}^n \\ s_{21}^n & s_{22}^n & s_{23}^n & s_{24}^n \\ s_{31}^n & s_{32}^n & s_{33}^n & s_{34}^n \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{41}$$

where s_{ij}^p denotes the (i, j)-element of $\mathbf{P}_p$. With $\lambda_p = s_{31}^p X + s_{32}^p Y + s_{33}^p Z + s_{34}^p$ and after some slight rearranging we obtain:

$$\underbrace{\begin{bmatrix} s_{31}^1 x_1 - s_{11}^1 & s_{32}^1 x_1 - s_{12}^1 & s_{33}^1 x_1 - s_{13}^1 \\ s_{31}^1 y_1 - s_{21}^1 & s_{32}^1 y_1 - s_{22}^1 & s_{33}^1 y_1 - s_{23}^1 \\ \vdots & \vdots & \vdots \\ s_{31}^n x_n - s_{11}^n & s_{32}^n x_n - s_{12}^n & s_{33}^n x_n - s_{13}^n \\ s_{31}^n y_n - s_{21}^n & s_{32}^n y_n - s_{22}^n & s_{33}^n y_n - s_{23}^n \end{bmatrix}}_{\mathbf{Q}} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \underbrace{\begin{bmatrix} s_{14}^1 - s_{34}^1 x_1 \\ s_{24}^1 - s_{34}^1 y_1 \\ \vdots \\ s_{14}^n - s_{34}^n x_n \\ s_{24}^n - s_{34}^n y_n \end{bmatrix}}_{\mathbf{r}}. \tag{42}$$

If $\text{rank}(\mathbf{Q}) = 3$, the least squares solution for $\hat{\mathbf{M}} = [\hat{X}\ \hat{Y}\ \hat{Z}]^\mathsf{T}$ is then given by:

$$\hat{\mathbf{M}} = [\mathbf{Q}^\mathsf{T}\mathbf{Q}]^{-1}\mathbf{Q}^\mathsf{T}\mathbf{r}. \tag{43}$$

5 Summary

In this paper a mathematical background of the multiple view geometry in relation to X-ray testing was presented. The key idea of the multiple view analysis is to gain more information about a test object by analysing multiple views taken at different viewpoints. Applications can be found in Part II of this paper [4].

Acknowledgment

This work was supported by the Departamento de Investigaciones Científicas y Tecnológicas of the Universidad de Santiago de Chile (grant DICYT 06-0119MQ). The author would like to thank the anonymous reviewers for their helpful comments.

References

[1] D. Mery, Th. Jaeger, and D. Filbert. A review of methods for automated recognition of casting defects. *INSIGHT, Journal of the British Institute of Non-Destructive Testing*, 44(7):428–436, 2002.

[2] Th. Jaeger, U. Heike, and K. Bavendiek. Experiences with an amorphous silicon array detector in an ADR application. In *International Computerized Tomography for Industrial Applications and Image Processing in Radiology, DGZfP Proceedings BB 67-CD*, pages 111–114, Berlin, March 15-17 1999.

[3] M. Purschke. IQI-sensitivity and applications of flat panel detectors and X-ray image intensifiers – a comparison. *INSIGHT, Journal of the British Institute of Non-Destructive Testing*, 44(10):628–630, 2002.

[4] D. Mery. Exploiting multiple view geometry in X-ray testing: Part II - Applications. *Materials Evaluation*, ??(??):???–???, 2003.

[5] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, Cambridge MA, London, 1993.

[6] R. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Trans. Robotics and Automation*, RA-3(4):323–344, 1987.

[7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[8] D. Mery and D. Filbert. Die Epipolargeometrie in der Röntgendurchleuchtungsprüfung: Grundlagen und Anwendung. *at - Automatisierungstechnik*, 48(12):588–596, 2000.

[9] D. Mery and D. Filbert. Epipolar geometry in radioscopic images. In *Proceedings of the International Symposium on Computerized Tomography for Industrial Applications and Image Processing in Radiology*, volume BB 67-CD, pages 181–187, Berlin, March 15-17 1999.

[10] D. Mery and D. Filbert. Automated flaw detection in aluminum castings based on the tracking of potential defects in a radioscopic image sequence. *IEEE Trans. Robotics and Automation*, 18(6):890–901, December 2002.

[11] Ch. Brack, H. Götte, F. Gossé, J. Moctezuma, M. Roth, and A. Schweikard. Towards accurate X-ray-camera calibration in computer-assisted robotic surgery. In *Proc. Int. Symp. Computer Assisted Radiology (CAR)*, pages 721–728, Paris, 1996.

[12] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distorsion models and accuracy evaluation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 4(10):965–980, 1992.

[13] R. Felix and B. Ramm. *Das Röntgenbild*. Georg Thieme Verlag, Stuttgart, New York, 3 edition, 1988.

[14] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[15] R. Hartley. Euclidean reconstruction from uncalibrated views. In *Proceeding of the DARPA-ESPRIT workshop on Applications of Invariants in Computer Vision*, pages 187–202, Azores, Portugal, Oct. 1993.

[16] Q.-T. Luong and O. Faugeras. Self calibration of a moving camera from point correspondences and fundamental matrices. *International Journal of Computer Vision*, 22(3):261–289, 1997.

[17] O. Faugeras and G. Toscani. The calibration problem for stereo. In *Proceedings IEEE Computer Vision and Pattern Recognition*, pages 15–20, 1986.

[18] H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[19] R. Hartley. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2):125–150, 1997.

[20] R. Hartley. Multilinear relationships between coordinates of corresponding image points and lines. In *Proceedings of the International Workshop on Computer Vision and Applied Geometry*, International Sophus Lie Center, Nordfjordeid, Norway, Aug. 1995.

[21] A. Heyden. A common framework for multiple view tensors. In *5th European Conference on Computer Vision (ECCV-98)*, pages 3–19, Jun. 1998.

[22] O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*. The MIT Press, Cambridge MA, London, 2001.

[23] R. Hartley. A linear method for reconstruction from lines and points. In *5th International Conference on Computer Vision (ICCV-95)*, pages 882–887, Cambridge, MA, 1995.

[24] O. Faugeras and T. Papadopulo. A nonlinear method for estimating the projective geometry of 3 views. In *6th International Conference on Computer Vision (ICCV-98)*, pages 477–484, Bombay, India, 1998.

[25] A. Heyden. Multiple view geometry using multifocal tensors. In *DSAGM*, Köpenhamn, 1999.

[26] A. Shashua and M. Werman. Trilinearity of three perspective views and its associated tensor. In *5th International Conference on Computer Vision (ICCV-95)*, Boston MA, Jun. 1995.

[27] D. Mery. *Automated Flaw Detection in Castings from Digital Radioscopic Image Sequences*. Verlag Dr. Köster, Berlin, 2001. (Ph.D. Thesis in German).

Domingo Mery (1965) received his Diploma (MS) degree in Electrical Engineering from the Technical University of Karlsruhe, Germany, in 1992, and his PhD degree with distinction at the Technical University of Berlin, in 2000. During 1993–1996, he was a Research Assistant at the Department of Electrical Engineering at the Catholic University of Chile. From 1996 to 2000, he was a Research Scientist at the Institute for Measurement and Automation Technology at the Technical University of Berlin with the collaboration of Philips and YXLON X-Ray International. From 1989 to 1992 he was a recipient of a Scholarship from the Konrad-Adenauer-Foundation, and from 1996 to 2000 of a Scholarship from the German Academic Exchange Service (DAAD). Since 2001 he is an Assistant Professor at the Departamento de Ingeniería Informática of the Universidad de Santiago de Chile. His research interests include image processing for automated inspection, X-ray imaging, computer vision and nondestructive testing and evaluation.

Table 1: Transformation between Euclidean and homogeneous coordinates.

Euclidean coordinates	$\leftrightarrow$	Homogeneous coordinates
	2 dimensions:	
(x, y)	$\rightarrow$	$\lambda(x, y, 1)$
$(x, y)/z$	$\leftarrow$	(x, y, z)
	3 dimensions:	
(X, Y, Z)	$\rightarrow$	$\lambda(X, Y, Z, 1)$
$(X, Y, Z)/W$	$\leftarrow$	(X, Y, Z, W)

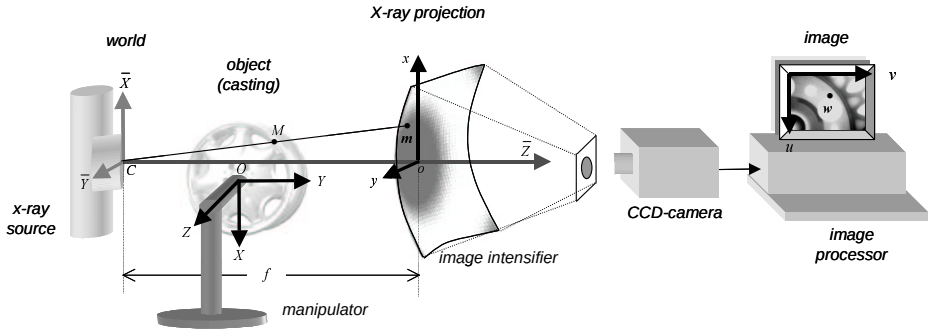

Figure 1: Diagram of an automated X-ray inspection system and its coordinate systems. World: $(\bar{X}, \bar{Y}, \bar{Z})$. Object: (X, Y, Z). X-ray projection: (x, y). Digital image: (u, v).

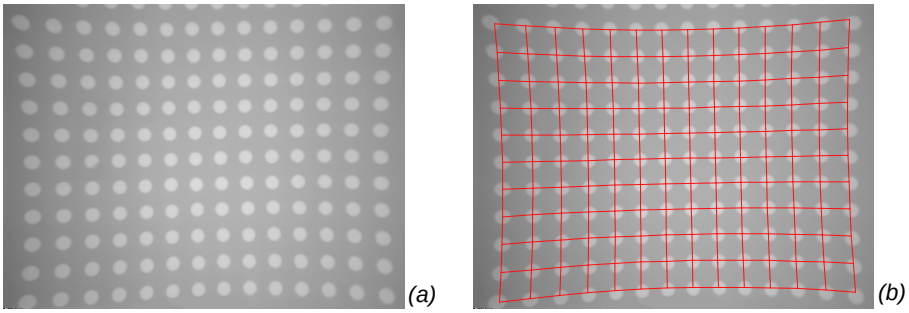

Figure 2: X-ray of the calibration plate and its modelled grid.

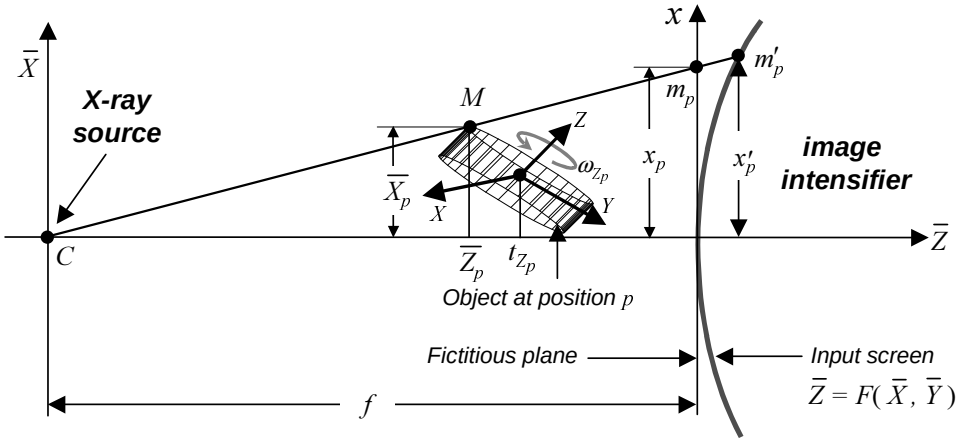

Figure 3: X-ray projective projection ($\bar{Y}$ and y axes are not shown).

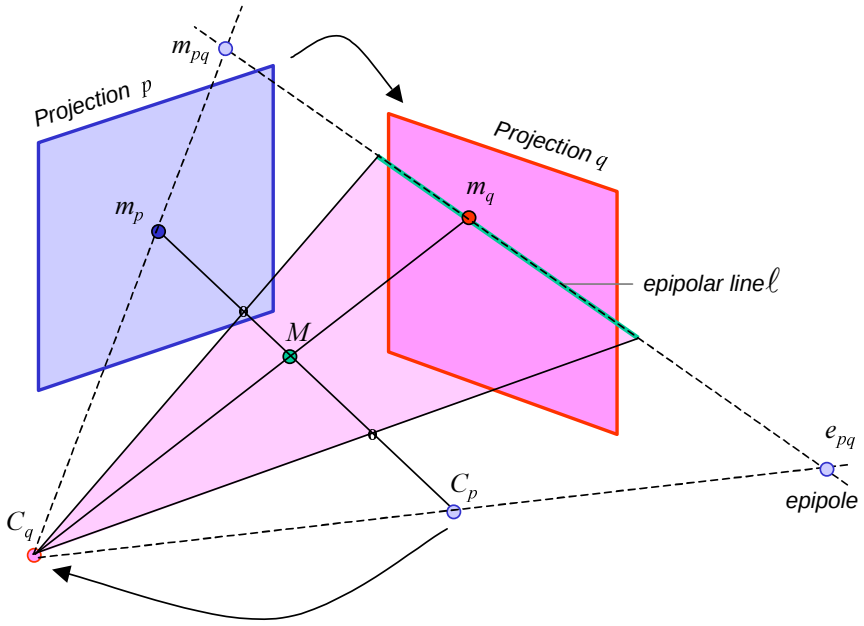

Figure 4: Epipolar geometry between two views.

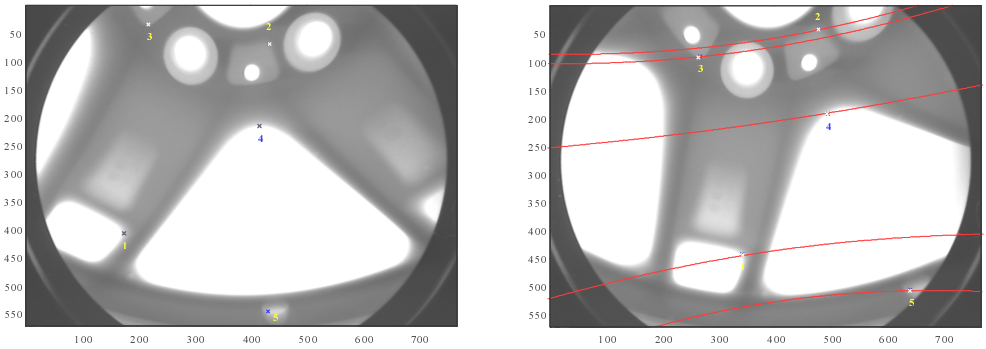

Figure 5: Epipolar lines using a hyperbolic model [8].

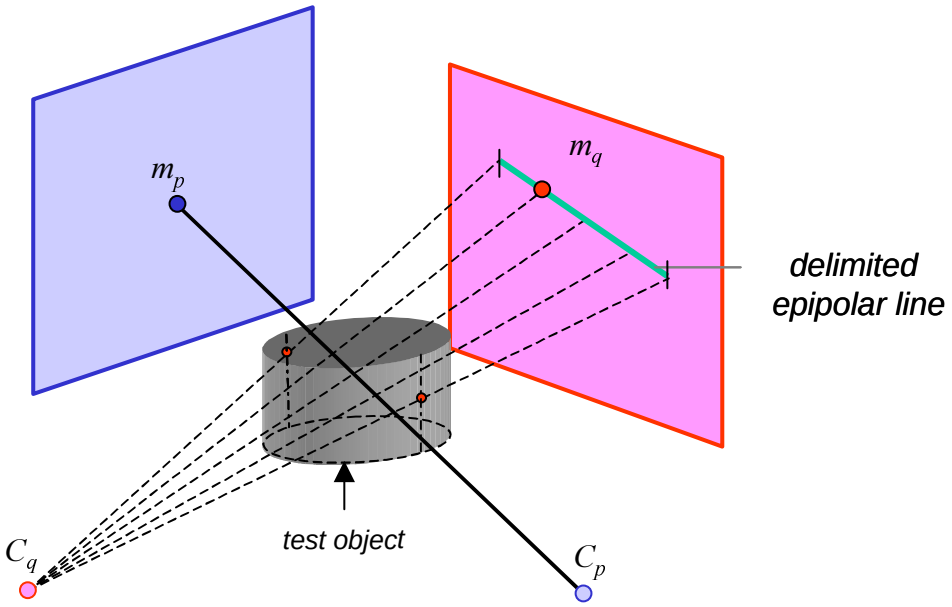

Figure 6: Epipolar geometry in two views using 3D information of the test object.

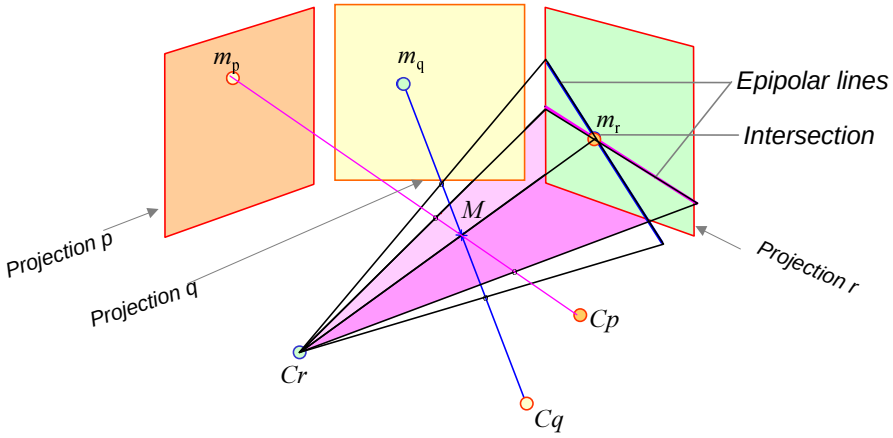

Figure 7: Epipolar geometry in three views.