AN INTRODUCTION TO THE RELATIVE ORIENTATION USING THE TRIFOCAL TENSOR

Camillo RESSL
University of Technology, Vienna, Austria
Institute of Photogrammetry and Remote Sensing
car@ipf.tuwien.ac.at

Working Group III/1

KEY WORDS: calibration, computer vision, mathematical models, orientation

ABSTRACT

About five years ago, in Computer Vision, a linear representation for the relative orientation of three images by means of the so-called Trifocal Tensor (TFT) was discovered using more parameters than necessary having no easily comprehensible geometric meaning compared to the notion of inner orientation (IOR) and exterior orientation (XOR). The relative orientation's fundamental condition of intersecting projection-rays for each homologous point-triple is described by four homogenous linear equations. The TFT also allows the usage of homologous image-lines for the relative orientation which is not possible for the relative orientation of two images. Each triple of homologous lines gives two linear equations. The TFT is made of 27 elements and so it can be computed linearly up to scale (since only homogenous equations are used) using ≥ 7 point-triples or ≥ 13 line-triples or combinations by means of a linear least-squares-adjustment minimising algebraic-error. Certain arrangements of the TFT's elements form matrices with interesting geometric properties, which can be used to compute the images IOR and XOR out of the tensor.

1 INTRODUCTION

The basic requirement for doing object reconstruction with a set of photographs is image orientation; i.e. the estimation of the XOR and maybe the IOR of all the photographs. For some of these tasks the reference to a global system of coordinates (*'absolute orientation'*) is either not necessary or done later: in other words, no or insufficient control features may be available. In such cases one works with the so-called *'relative orientation'*. This is the alignment of at least two images in such a way that homologous projection rays intersect each other in a point in space.

The relative orientation of images is determined using only the observed image coordinates which are subject to accidental errors. To decrease the error's disturbing influence on the estimated unknowns an adjustment is done during which the sum of squares of the errors is minimized. For this, two statistically well-founded models exist: the Gauss-Markoff-model, in which each observation can be expressed in terms of the unknowns (aka '*adjustment by indirect observations*'), and the Gauss-Helmert-model, in which only combinations of the observations can be expressed in terms of the unknowns (aka '*general case of least squares adjustment*').

The unknowns which are estimated in these two models are distinguished by the following properties: They are unbiased and have least variance ('*best unbiased estimation*'), the sum of the squared discrepancies is minimized ('*least-squares-estimation*') and in the case of observations having a normal distribution they are a '*maximum-likelihood-estimation*';[Koch 1987]. In general, one works with a least-squares-estimation. However, it is important to point out that such a least-squares-estimation is a best unbiased estimation only if the sum of the squares of the original observations' errors (so-called '*measurement-error*' or '*reprojection-error*') is minimized and not the sum of squares of some other quantities (so-called '*algebraic-error*').

For such an adjustment, however, linear equations are required. Since the equations of the central projection are non-linear, they have to be linearized, and for this approximate values of the XOR (and under certain circumstances also for the IOR) of the photographs are necessary, but in many cases the determination of these approximate values is quite tedious.

In the relatively young discipline of computer vision central perspective images are also worked with. Due to the highly non-linear character of the central-projection formulated in terms of XOR and IOR in computer vision a linear representation for the central perspective relation between object and image is aimed for. This linear representation is

achieved by projective geometry but at the price of using more parameters than necessary, which have no easily comprehensible geometric meaning compared to the notion of IOR and XOR. The linear representation of the central projection is obtained by means of a 3x4 matrix with 11 degrees of freedom (DOF) – the '*projection matrix*' (see section 2.1). Furthermore, certain indexed systems of numbers (so-called tensors) are used in computer vision, which describe the relative orientation of 2, 3 and 4 images in a linear manner. For five and more images no such alternative formulization exists. Using the so-called '*essential matrix*' resp. '*fundamental matrix*' one gets a linear representation of the relative orientation of two calibrated resp. uncalibrated images (see section 2.2). The so-called '*trifocal tensor*' T_i^{jk} (TFT) enables a linear form of the relative orientation of three uncalibrated images. Moreover, with the so-called '*quadrifocal tensor'* the relative orientation of four uncalibrated images can be presented linearly; e.g. [Hartley 1998]. Within this paper a closer look at the trifocal tensor and the relative orientation of three images will be given.

2 BASICS

2.1 The central projection using homogenous coordinates

Using homogenous coordinates it is possible to write the central projection of an image ψ in a very compact way. If the XOR of ψ is given by the image's projection center $\mathbf{O}_\psi$ and rotation matrix $\mathbf{R}$ (from the image system to the global system of coordinates), and if the IOR of the image is given by the principal point (x_0, y_0), the principal distance f and two parameters $(\alpha\ \beta)$ modeling affine image deformations, then the central perspective image-point $\mathbf{p}^T = (x\ y\ 1)$ – as a homogenous vector – of an object point $\mathbf{P}^T = (X\ Y\ Z)$ may be computed by the following product of matrices:

$$\mathbf{p} \sim \mathbf{C}^{-1} \cdot \mathbf{R}^T \cdot [\mathbf{E}_{3\times3}, -\mathbf{O}_\psi] \cdot \begin{bmatrix} \mathbf{P} \\ 1 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 1 & \alpha & -x_0 \\ 0 & \beta & -y_0 \\ 0 & 0 & -f \end{bmatrix} \qquad \mathbf{E}_{3\times3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.1}$$

'~' symbolizes that the left and right side in (2.1) are equal only up to scale. Using these matrices and vector $\mathbf{p}$ it is possible to write the projection ray $\mathbf{r}$ from the center of projection to the object point $\mathbf{P}$ in a compact way: $\mathbf{r} = \mathbf{R} \cdot \mathbf{C}_1 \cdot \mathbf{p}$. If a straight line *m* in the object space is given, then, in general, its projection in the image will also be a straight line $\lambda_\mathbf{m}$. Due to the point-line-duality in the 2-dimensional projective space it is possible to identify the straight line $\lambda_\mathbf{m}$ by a homogenous vector $\lambda_\mathbf{m} = (a\ b\ c)^T$ (up to scale). A reasonable way to define the scale of $\lambda_\mathbf{m}$ would be to set $a^2 + b^2 = 1$. In this case *c* would be the orthogonal Euclidean distance of the image coordinate system's origin to $\lambda_\mathbf{m}$. An image-point $\mathbf{p}$ is sited on $\lambda_\mathbf{m}$ if it holds: $\mathbf{p}^T \cdot \lambda_\mathbf{m} = 0$. Furthermore, given the image line $\lambda_\mathbf{m}$ and the elements of IOR and XOR it is easy to determine the normal-vector $\mathbf{n}_\varepsilon$ of the projection plane ε going through $\mathbf{O}_\psi$, λ_m and *m* by: $\mathbf{n}_\varepsilon = \mathbf{R} \cdot \mathbf{C}^{-T} \cdot \lambda_\mathbf{m}$

Note1: A linear mapping of points to points within the 2-dimensional projective space is termed '*collineation*' and a linear mapping of points to lines within the 2-dimensional projective space is termed '*correlation*'.

Note2: The matrix-multiplication $\mathbf{C^{-1} \cdot R \cdot [E_{3\times3}, -O_\psi]}$ would result in a 3×4-matrix – the projection-matrix – having 11 DOF due to the scale ambiguity.

2.2 The relative orientation of two images

The relative orientation of two images ψ_1 and ψ_2 (with IORs $\mathbf{C_1}$ and $\mathbf{C_2}$ and with projection centers $\mathbf{O_1}$ and $\mathbf{O_2}$ and attitudes $\mathbf{R_1}$ and $\mathbf{R_2}$) is obtained by the intersection of the projection rays of homologous image-points $\mathbf{p_1}$ and $\mathbf{p_2}$; i.e. the base-vector $\mathbf{b} = \mathbf{O_2} - \mathbf{O_1}$ and the two projection rays are coplanar. Using the results of section 2.1 we can formulate this relation in an elegant manner using a quadratic form; e.g. [Niini 1994]:

$$(\mathbf{R_1} \cdot \mathbf{C_1} \cdot \mathbf{p_1})^T \cdot (\mathbf{b} \times (\mathbf{R_2} \cdot \mathbf{C_2} \cdot \mathbf{p_2})) = 0 \quad \Rightarrow \quad \mathbf{p_1^T} \cdot \underbrace{\mathbf{C_1^T} \cdot \mathbf{R_1^T} \cdot [\mathbf{b}]_\times \cdot \mathbf{R_2} \cdot \mathbf{C_2}}_{\mathbf{F_{12}^T}} \cdot \mathbf{p_2} = \underline{\underline{\mathbf{p_1^T} \cdot \mathbf{F_{12}^T} \cdot \mathbf{p_2} = 0}} \tag{2.2}$$

The cross-product was solved by:

$$[\mathbf{b}]_\times = \begin{bmatrix} 0 & -b_Z & b_Y \\ b_Z & 0 & -b_X \\ -b_Y & b_X & 0 \end{bmatrix} \tag{2.3}$$

This property of intersection is independent of the origin and attitude of the object-coordinate-system. As a consequence, the latter may be fixed, e.g., in the following way: Its origin coincides with the projection center of the first image and the attitude of the axes of this photograph's image-coordinate-system fixes the orientation of the object-

coordinate-system's axes: $\mathbf{O_1} = \mathbf{0}$, $\mathbf{b} = \mathbf{O_2}$, $\mathbf{R_1} = \mathbf{E_{3\times3}}$ (the so-called '*relative orientation by successive images*'). The matrix $\mathbf{F_{12}}$ represents an alternative formulization of the relative orientation of two images in a linear way. In the computer vision community $\mathbf{F_{12}}$ is termed '*essential matrix*' (e.g. [Longuet-Higgins 1981]) in the case of calibrated images and '*fundamental matrix*' in the case of uncalibrated images (e.g. [Luong, Faugeras 1996]) whereas in the photogrammetric community $\mathbf{F_{12}}$ is always termed '*correlation matrix*' (e.g. [Brandstätter 1991], [Niini 1994]).

The latter name is derived from $\mathbf{F_{12}}$'s property that $\mathbf{p_1}$ resp. $\mathbf{p_2}$ is mapped by $\mathbf{F_{12}}$ to the corresponding epipolar-line in the 2nd image ($\mathbf{F_{12}}\cdot\mathbf{p_1}$) resp. 1st image ($\mathbf{F^T_{12}}\cdot\mathbf{p_2}$) on which the homologous partner $\mathbf{p_2}$ resp. $\mathbf{p_1}$ is constrained to lie (called '*epipolar constraint*'). So by means of $\mathbf{F_{12}}$ not only the relative orientation of 2 images is described in a linear way, but it also allows a compact representation of the epipolar geometry inherent in two images. As a consequence, the epipoles and epipolar-lines can easily be computed. The epipoles $\mathbf{v_{12}}$ and $\mathbf{v_{21}}$ are the left and right kernels of $\mathbf{F_{12}}$; i.e. $\mathbf{F_{12}}\cdot\mathbf{v_{12}} = 0$ and $\mathbf{F^T_{12}}\cdot\mathbf{v_{21}} = 0$. $\mathbf{F_{12}}$ is also used for the so-called '*epipolar transfer*', i.e. given 3 images and the fundamental matrices $\mathbf{F_{12}}$, $\mathbf{F_{13}}$, $\mathbf{F_{23}}$ and the homologous image-points in 2 images, the homologous image-point in the 3rd image can be computed by means of intersecting epipolar lines. But it should be noted that this epipolar transfer fails if the corresponding object-point and the three projection-centers lie in one common plane. All these properties may be the reason why $\mathbf{F_{12}}$ is also termed '*essential*' resp. '*fundamental*'.

2.3 The homography induced by a plane

A '*homography*' (according to [Shashua, Werman 1995] '*a projective transformation of planes*') is a mapping (collineation) of points from one image ψ_1 to the points of another image ψ_2. By means of a 3×3 matrix $\mathbf{H_{12,\sigma}}$ the homography can be expressed in the following way:

$$\mathbf{p_2} \sim \mathbf{H_{12,\sigma}}\cdot\mathbf{p_1} \tag{2.4}$$

The points $\mathbf{p_1}$ and $\mathbf{p_2}$ which are related in this way by $\mathbf{H_{12,\sigma}}$ are the image-points of a point in 3d-space. The 3d-space-points of all pairs of image-points which satisfy (2.4) lie in one common plane – the homography-plane σ. One may say: "$\mathbf{H_{12,\sigma}}$ is a homography from image ψ_1 to image ψ_2 induced by the plane σ". Homographies can be used e.g. for the detection of obstacles sited on a plane on which a robot is moving. If the XOR and IOR of the two images ψ_1 and ψ_2 are given in the same way as in section 2.1 and if the plane σ is given by $\mathbf{n_\sigma^T}\cdot\mathbf{P} = d$; with $\mathbf{n_\sigma}$ being the plane's normal-vector, then $\mathbf{H_{12,\sigma}}$ has the following structure:

$$\mathbf{H_{12,\sigma}} = \mathbf{C_2^{-1}}\cdot\mathbf{R_2^T}\left[\mathbf{E_{3\times3}} - \frac{1}{\left(d - \mathbf{n_\sigma^T}\cdot\mathbf{O_1}\right)}\left(\mathbf{O_2} - \mathbf{O_1}\right)\cdot\mathbf{n_\sigma^T}\right]\cdot\mathbf{R_1}\cdot\mathbf{C_1} \tag{2.5}$$

The homography $\mathbf{H_{21,\sigma}}$ due to the same plane σ but from image ψ_2 to image ψ_1 looks the same, except for exchanging the indices 1 and 2. Generally rank($\mathbf{H_{12,\sigma}}$) will be 3. Depending on rank($\mathbf{H_{12,\sigma}}$) the following relations between $\mathbf{H_{12,\sigma}}$ and the epipoles $\mathbf{v_{12}}$ and $\mathbf{v_{21}}$ hold; [Ressl 1997]:

$rank(\mathbf{H_{12,\sigma}}) = 3$:	$rank(\mathbf{H_{12,\sigma}}) = 2$:	$rank(\mathbf{H_{12,\sigma}}) = 1$:
$\mathbf{v_{21}} \sim \mathbf{H_{12,\sigma}}\cdot\mathbf{v_{12}}$	$\mathbf{O_2} \in \sigma \Leftrightarrow \left(d - \mathbf{n_\sigma^T}\cdot\mathbf{O_2}\right) = 0$ $\mathbf{0} = \mathbf{H_{12,\sigma}}\cdot\mathbf{v_{12}}$ $\mathbf{H_{12,\sigma}}\cdot\mathbf{p_1} \in s_2 : \forall\,\mathbf{p_1} \neq \mathbf{v_{12}}$ *with* $s_2 : \mathbf{n_\sigma^T}\cdot\mathbf{R_2}\cdot\mathbf{C_2}\cdot\mathbf{p_2} = 0$ (*a line in* ψ_2)	$\mathbf{O_1} \in \sigma \Leftrightarrow \left(d - \mathbf{n_\sigma^T}\cdot\mathbf{O_1}\right) = 0$ $\mathbf{H_{12,\sigma}}\cdot\mathbf{p_1} \sim \begin{cases} \mathbf{v_{21}} & : \forall\,\mathbf{p_1} \notin s_1 \\ \mathbf{0} & : \forall\,\mathbf{p_1} \in s_1 \end{cases}$ *with* $s_1 : \mathbf{n_\sigma^T}\cdot\mathbf{R_1}\cdot\mathbf{C_1}\cdot\mathbf{p_1} = 0$ (*a line in* ψ_1)

Finally there is also a relation between $\mathbf{H_{12,\sigma}}$ (rank >1)and the fundamental-matrix $\mathbf{F_{12}}$ of the two images:

$$\mathbf{F_{12}} \sim [\mathbf{v_{21}}]_\times\cdot\mathbf{H_{12,\sigma}} \tag{2.6}$$

2.4 A few basics of tensor calculus

A tensor is an indexed system of numbers. There are two kinds of indices: sub-indices are called '*co-variant*' and super-indices '*contra-variant*'. A tensor with *contra-variant valence p* and *co-variant valence q* has n^{p+q} components with n being the dimension of the underlying vector-space; i.e. each index runs from 1 to n. Using these indices and Einstein's

convention of summation certain mathematical relations can be expressed in a very efficient way. This convention says that a sum is made of all the same indices appearing as co- and contra-variant. So, for example the scalar product $s(\boldsymbol{x},\boldsymbol{y}) = \mathbf{x}^{\mathbf{T}}\cdot\mathbf{y}$ of two vectors $\mathbf{x} = (x^1\ x^2\ x^3)^T$ and $\mathbf{y} = (y^1\ y^2\ y^3)^T$ can be written in a shorter way as: $s(\boldsymbol{x},\boldsymbol{y}) = x_i y^i$. The product $\mathbf{A}\cdot\mathbf{B} = \mathbf{C}$ of two matrices **A** and **B** can be written as $A^i_j \cdot B^j_k = C^i_k$. The contra-variant indices relate to the rows and the co-variant ones to the columns. As it can be seen the index responsible for the summation disappears (*i* resp. *j* in the examples above). Such indices are termed *saturated* (or *dummy*) while the remaining ones are termed *free* indices. Using this indexation and the convention of summation the following expression of matrices may be simplified: $\mathbf{A}\cdot\mathbf{E}\cdot\mathbf{B} + \mathbf{C}\cdot\mathbf{E}\cdot\mathbf{D} = \mathbf{F} \Rightarrow A^i_j \cdot E^j_k \cdot B^k_m + C^i_j \cdot E^j_k \cdot D^k_m = E^j_k \cdot (A^i_j \cdot B^k_m + C^i_j \cdot D^k_m) = F^i_m$. The indexation and the convention of summation are everything of tensor calculus that is needed for the following.

3 A HISTORICAL REVIEW

The existence of dependencies among 3 images of an object has been known in photogrammetry since, e.g., [Rinner, Burkhardt 1972] (trilinear image-fields resp. trilinear relations). The term '*trilinear*' means, in this case, that a homologous line in the first and second image is related to a line in the third image. These trilinear relations are used, e.g., to transfer image contents (image-pair of maximum image-content) or to create an object's ground plan by means of two images. Despite those and other interesting properties of these trilinear relations, they remained quite unused in photogrammetry, maybe due to the lack of a compact mathematical formulization.

In computer vision the first to discover redundancies within the contents of 3 images were [Spetsakis, Aloimonos 1990]. They found three relations between homologous points and one relation between homologous lines in three images expressed in terms of 27 coefficients, arranged in three 3×3-matrices. [Shashua 1995] showed that as a matter of fact these 27 coefficients and one homologous triple of points in three views form together nine linear equations (four of them being independent), which he called '*trilinearities*', since they consist of products of three image coordinates and one of the 27 coefficients. Furthermore [Hartley 1994] showed that the same 27 coefficients and a homologous triple of lines actually create two equations. He also proposed for this set of 3x3x3 coefficients the term '*trifocal tensor*', nevertheless the tensor is sometimes also referenced as '*trilinear tensor*'.

4 THE TRIFOCAL TENSOR AND ITS OBSERVATION EQUATIONS

Given are three images ψ_1, ψ_2, ψ_3. In view of a relative orientation the first image's projection center $\mathbf{O_1}$ is set to $\mathbf{0}$. So the central projection of these three images is given by:

$$\begin{aligned}
&\mathbf{p_1} = \mathbf{C_1^{-1}}\cdot\mathbf{R_1^T}\cdot[\mathbf{E_{3\times3}},\mathbf{0}]\cdot\begin{bmatrix}\mathbf{P}\\1\end{bmatrix}\\
&\mathbf{p_2} = \mathbf{C_2^{-1}}\cdot\mathbf{R_2^T}\cdot[\mathbf{E_{3\times3}},-\mathbf{O_2}]\cdot\begin{bmatrix}\mathbf{P}\\1\end{bmatrix} \qquad \mathbf{A} = \mathbf{C_2^{-1}}\cdot\mathbf{R_2^T}\cdot\mathbf{R_1}\cdot\mathbf{C_1} \qquad \mathbf{v_{21}} = -\mathbf{C_2^{-1}}\cdot\mathbf{R_2^T}\cdot\mathbf{O_2}\\
&\mathbf{p_3} = \mathbf{C_3^{-1}}\cdot\mathbf{R_3^T}\cdot[\mathbf{E_{3\times3}},-\mathbf{O_3}]\cdot\begin{bmatrix}\mathbf{P}\\1\end{bmatrix} \qquad \mathbf{B} = \mathbf{C_3^{-1}}\cdot\mathbf{R_3^T}\cdot\mathbf{R_1}\cdot\mathbf{C_1} \qquad \mathbf{v_{31}} = -\mathbf{C_3^{-1}}\cdot\mathbf{R_3^T}\cdot\mathbf{O_3}
\end{aligned} \tag{4.1}$$

If the images $\lambda_1, \lambda_2, \lambda_3$ of a 3d-straight-line *m* (not lying in the epipolar-plane of $\mathbf{O_2}$ and $\mathbf{O_3}$) are given then it can be shown that the following relation holds:

$$\boldsymbol{\lambda}_1^{\mathbf{T}} \sim -\left(\boldsymbol{\lambda}_3^{\mathbf{T}}\cdot\mathbf{v}_{31}\right)\cdot\boldsymbol{\lambda}_2^{\mathbf{T}}\cdot\mathbf{A} + \left(\boldsymbol{\lambda}_2^{\mathbf{T}}\cdot\mathbf{v}_{21}\right)\cdot\boldsymbol{\lambda}_3^{\mathbf{T}}\cdot\mathbf{B} \tag{4.2}$$

So using (4.2) the mapping of *m* in image ψ_1 can be computed using the images of *m* in ψ_2 and ψ_3 and the orientation elements **A**, **B**, $\mathbf{v_{21}}$, $\mathbf{v_{31}}$. If the products and sums in (4.2) are replaced by tensorial-expressions we get the following three-lines-relation (the indices *i*, *j*, *k* running from 1 to 3 are put in brackets for better distinction):

$$\lambda_{1(i)} \sim -\lambda_{3(k)}\cdot v_{31}^{(k)}\cdot\lambda_{2(j)}\cdot A_{(i)}^{(j)} + \lambda_{2(j)}\cdot v_{21}^{(j)}\cdot\lambda_{3(k)}\cdot B_{(i)}^{(k)} = \lambda_{2(j)}\cdot\lambda_{3(k)}\cdot\underbrace{\left(v_{21}^{(j)}\cdot B_{(i)}^{(k)} - v_{31}^{(k)}\cdot A_{(i)}^{(j)}\right)}_{T_i^{j,k}} \Rightarrow \underline{\underline{\lambda_{1(i)} \sim \lambda_{2(j)}\cdot\lambda_{3(k)}\cdot T_i^{j,k}}} \tag{4.3}$$

T_i^{jk} is the trifocal tensor which is not unique, because it depends on the choice of the 1st image (this is expressed in (4.3) in such a way that only λ_1 can be predicted by λ_2 and λ_3 – no way for a prediction of λ_2 by λ_1 and λ_3 or of λ_3 by λ_1 and λ_2). If one multiplies this three-lines-relation (4.3) by $\mathbf{p_1} = (x_1 \quad y_1 \quad 1)^T = p_1^{(i)}$ – a point in image ψ_1 lying on λ_1 – one gets:

$$p_1^{(i)}\lambda_{1(i)}=0=p_1^{(i)}\cdot\lambda_{2(j)}\cdot\lambda_{3(k)}\cdot T_i^{j,k} \tag{4.4}$$

This is one single point-line-line-equation. If instead of the homologous lines λ_2 and λ_3 the homologous points $\mathbf{p_2}=(x_2 \;\; y_2 \;\; 1)^T$ and $\mathbf{p_3}=(x_3 \;\; y_3 \;\; 1)^T$ are known – so $\{\mathbf{p_1} \;\; \mathbf{p_2} \;\; \mathbf{p_3}\}$ forming a homologous point-triple – one can use these image-points to create fictitious lines in each of the two images (ψ_2, ψ_3). With h = (2,3) these three lines are the line $\boldsymbol{\mu}_{\mathbf{x,h}}=(0 \;\; 1 \;\; -y_h)^T$ going through $\mathbf{p_h}$ parallel to the x-axis of image ψ_h, the line $\boldsymbol{\mu}_{\mathbf{y,h}}=(1 \;\; 0 \;\; -x_h)^T$ going through $\mathbf{p_h}$ parallel to the y-axis of image ψ_h and the line $\boldsymbol{\mu}_{\mathbf{0,h}}=(y_h \;\; -x_h \;\; 0)^T$ connecting the origin of the image-system of image ψ_h and the image-point $\mathbf{p_h}$. Using these three lines one has 3×3 = 9 possibilities to choose λ_2 resp λ_3 out of the set $\{\mu_{x,2} \;\; \mu_{y,2} \;\; \mu_{0,2}\}$ resp. $\{\mu_{x,3} \;\; \mu_{y,3} \;\; \mu_{0,3}\}$ to form equation (4.4). These nine equations are the nine trilinearities reported in [Shashua 1995]. Since only four of them are linearly independent one may erase $\mu_{0,h}$ and combine the other two to the matrices $\mathbf{L_h}=\begin{bmatrix}1 & 0 & -x_h\\ 0 & 1 & -y_h\end{bmatrix}$. Then, using again tensorial-expressions we can write four independent trilinearities (three-points-relations) in the following form, which are valid as long as no image-point $\{\mathbf{p_2} \;\; \mathbf{p_3}\}$ is sited on the line of infinity:

$$p_1^{(i)}\cdot L_{2(j)}^{(l)}\cdot L_{3(k)}^{(m)}\cdot T_i^{j,k}=0^{(l,m)} \qquad \begin{matrix} l,m=\{1,2\} \\ i,j,k=1(1)3 \end{matrix} \tag{4.5}$$

The fourth possible relation within three images would be a point-point-line-relation. Given the homologous point $\mathbf{p_1}$ and the homologous line λ_2 resp. λ_3 then the homologous point $\mathbf{p_2}$ resp. $\mathbf{p_3}$ can be computed by:

$$p_2^{(j)}\sim p_1^{(i)}\cdot\lambda_{3(k)}\cdot T_i^{j,k} \quad resp. \quad p_3^{(k)}\sim p_1^{(i)}\cdot\lambda_{2(j)}\cdot T_i^{j,k} \tag{4.6}$$

The trifocal tensor includes all the projective geometric constraints inherent in three views, so it plays the same role for three views as the fundamental matrix plays for two. If one has homologous points or lines in two views, the location of the homologous partner in the third view can be computed by means of the tensor using equations (4.6) resp. (4.2). This '*tensorial-transfer*' of points works even in the case when the epipolar-transfer of points by means of fundamental matrices fails. The tensorial-transfer of lines, however, fails for object-lines lying in the epipolar-plane of $\mathbf{O_2}$ and $\mathbf{O_3}$.

Although the tensor depends on the XOR and IOR of the three views, it can be computed from image correspondences alone. So if one wants to compute the trifocal tensor for three given images, it holds: Every homologous triple of points resp. every homologous triple of lines gives four resp. two independent equations expressed linearly in the elements of the tensor; i.e. the equations (4.5) resp. (4.3). The trifocal tensor allows the use of homologous lines for the relative orientation for the first time; this was not possible with the fundamental matrix.

Since the tensor is made of 3×3×3 = 27 elements it may be computed given a sufficient number of point-triples (≥ 7) or line-triples (≥ 13) or combinations using a least-squares-adjustment by minimising the homogenous equations' right side. Since these equations (4.5) resp. (4.3) are homogenous the trifocal tensor is determined only up to scale (same property for the fundamental matrix) and so this scale has to be chosen; e.g. by $\|T_i^{jk}\| = 1$, which leads to an eigen-value problem [Hartley 1994]. For such a linear solution no approximate tensor is needed. One must be aware, however, of the fact that such a linear solution is obtained by minimising non-meaningful quantities whose minimization is justified neither geometrically nor stochastically, since not all of the point observations are treated the same way, i.e. the observations' weights become dependent on the point location in the images (i.e. minimising '*algebraic-error*' instead of '*measurement-error*').

Besides the fact of minimising algebraic-error, this linear solution has also the disadvantage that it is totally over-parameterised because it is computed with 26 unknowns although the relative orientation in the case of three calibrated images has only 11 DOF and in the case of three uncalibrated images 18 DOF. So we see that the tensor's elements should meet with 16 resp. 9 constraints (one of these constraints is the tensor's scale which has to be fixed in advance). So, this unconstrained linear solution minimising algebraic-error will have a strong bias depending on the image-noise (and the image-distortion, which are not included in the TFT-modelling). By minimising measurement-error without considering the mentioned constraints the situation gets slightly better. According to [Torr, Zisserman 1997] these constraints have been investigated but are not as yet thoroughly understood. In [Papadopoulo, Faugeras 1998] 12 constraints can be found; i.e. some of them should be dependent. All of these constraints are non-linear and are partly expressed in terms of the epipoles.

To compute a tensor that meets with these constraints but avoiding to explicitly include them in the computation one possibility is to introduce a new parameterisation for the tensor having exactly 18 coefficients. One such method is presented by [Torr, Zisserman 1997]: By choosing $\mathbf{E_{3\times3}}$ for the IOR-matrix $\mathbf{C_1}$ – which makes no difference for the projective relations – it is possible to compute the tensor by six homologous point triples across three views. Of the 36 image-coordinates of these triples, convenient 18 coordinates are kept fixed. In this way a consistent and minimal parameterisation of the tensor is achieved. The unknowns themselves are obtained as (up to 3) solutions of a cubic equation. Due to this fixing of erroneous observations in the images one may entertain suspicion that errors in the calculated tensor may be induced, furthermore no correct minimisation of the measurement-errors of all observations is possible. And as it is shown by the results in [Torr, Zisserman 1997] the standard deviation depends on the choice of the 6 points resp. the fixed 18 coordinates.

Another consistent and minimal parameterisation is proposed by [Papadopoulo, Faugeras 1998]. There 18 special quantities are selected as unknowns, parameterising the epipoles and some parts of the tensor. The solution for the selected unknowns is unique and is achieved by minimising measurement-error. This method depends on approximate values; e.g. obtained by the eigen-value algorithm mentioned. Furthermore that parameterisation itself is not unique; i.e. depending on the configuration of the three images other 18 quantities have to be chosen.

The photogrammetric standard case of known IOR is not dealt with in the papers of the computer vision community. Using calibrated images even 15 constraints must hold among the 27 elements of the tensor. Their form is still not known, either.

5 THE HOMOGRAPHIES AND CORRELATIONS OF THE TENSOR

One can imagine the trifocal tensor T_i^{jk} formed as a 3×3×3 cube of numbers and the cube's edges related to the indices i, j, k. If we keep one index fixed we slice a 3×3 matrix out of the tensor. Since we have three indices we get three different kinds of matrices – different also in their geometrical meaning. If we keep the *i*-index fixed as $i = t = \{1,2,3\}$, we get the following matrix $\mathbf{I_t}$ ($\mathbf{e_t}$ being the t[th] column of $\mathbf{E_{3\times3}}$):

$$\mathbf{I_t} = \mathbf{v_{21}} \cdot \mathbf{e_t^T} \cdot \mathbf{B^T} - \mathbf{A} \cdot \mathbf{e_t^T} \cdot \mathbf{v_{31}^T} \tag{4.1}$$

$\mathbf{I_t}$ describes a linear mapping (correlation) of the lines λ_3 in image ψ_3 to the points $\mathbf{p_2}$ in the image ψ_2. Such a point $\mathbf{p_2}$ in ψ_2 is the image-point of the intersecting-point of the projection-plane due to λ_3 and the projection-ray $\mathbf{R_1} \cdot \mathbf{C_1} \cdot \mathbf{e_t}$ of image ψ_1. Therefore, all mapped image-points $\mathbf{p_2}$ lie on one common line $\mathbf{l_t}$ in image ψ_2 – the epipolar-line of this particular projection-ray. The line $\mathbf{l_t}$ can be computed by:

$$\mathbf{I_t^T} \cdot \mathbf{l_t} = \mathbf{0} \qquad \textit{or by} \quad \mathbf{l_t} \sim [\mathbf{v_{21}}]_\times \cdot \mathbf{A} \cdot \mathbf{e_t} \sim \mathbf{F_{12}} \cdot \mathbf{e_t} \tag{4.2}$$

Since $\mathbf{l_t}$ is the left kernel of $\mathbf{I_t}$ it is also interesting to look at the right kernel $\mathbf{r_t}$ of $\mathbf{I_t}$ which is the epipolar-line in ψ_3 of that particular projection-ray:

$$\mathbf{I_t} \cdot \mathbf{r_t} = \mathbf{0} \qquad \textit{or by} \quad \mathbf{r_t} \sim [\mathbf{v_{31}}]_\times \cdot \mathbf{B} \cdot \mathbf{e_t} \sim \mathbf{F_{13}} \cdot \mathbf{e_t} \tag{4.3}$$

If all three $\mathbf{r_t}$ and $\mathbf{l_t}$ are computed and arranged as the rows of the matrices $\mathbf{R_\times}$ and $\mathbf{L_\times}$ then the epipols $\mathbf{v_{21}}$ and $\mathbf{v_{31}}$ can be computed quite easily:

$$\mathbf{L_\times} \cdot \mathbf{v_{21}} = \mathbf{0} \qquad \textit{and} \qquad \mathbf{R_\times} \cdot \mathbf{v_{31}} = \mathbf{0} \tag{4.4}$$

The columns of $\mathbf{I_t}$ are linear-combinations of the two vectors $\mathbf{v_{21}}$ and $\mathbf{A} \cdot \mathbf{e_t}$, so the rank of $\mathbf{I_t}$ will be 2 in general. These three $\mathbf{I_t}$-matrices are the basic element for [Papadopoulo, Faugeras 1998] to find their consistent and minimal parameterisation. The 12 constraints they found are: rank($\mathbf{R_X}$) = 2, rank($\mathbf{L_X}$) = 2 and $\det\left(\sum_{t=1}^{3} \gamma_t \cdot \mathbf{I_t}\right) = 0 \ \forall \gamma_t$ (which are 10 constraints).

If we keep the *j*-index fixed as $j = t = \{1,2,3\}$, we get the following matrix $\mathbf{J_t}$ ($\mathbf{e_t}$ being the t[th] column of $\mathbf{E_{3\times3}}$):

$$\mathbf{J_t} = v_{21}^{(t)} \cdot \mathbf{B} - \mathbf{v_{31}} \cdot \mathbf{e_t^T} \cdot \mathbf{A} \tag{4.5}$$

$\mathbf{J_t}$ is a homography-matrix from image ψ_1 to image ψ_3 due to the plane α_t whose position and attitude is determined by image ψ_2; in detail α_t goes through $\mathbf{O_2}$ with $\mathbf{n_{\alpha t}} = \mathbf{R_2}\cdot\mathbf{C_2}^{-\mathbf{T}}\cdot\mathbf{e_t}$ and $d_{\alpha t} = -v_{21}^{(t)}$. Similarly if we keep the *k*-index fixed as $k = t = \{1,2,3\}$, we get the following matrix $\mathbf{K_t}$ ($\mathbf{e_t}$ being the t^{th} column of $\mathbf{E_{3\times3}}$):

$$\mathbf{K_t} = \mathbf{v_{21}}\cdot\mathbf{e_t^T}\cdot\mathbf{B} - v_{31}^{(t)}\cdot\mathbf{A} \tag{4.6}$$

$\mathbf{K_t}$ is a homography-matrix from image ψ_1 to image ψ_2 due to the plane β_t whose position and attitude is determined by image ψ_3; in detail β_t goes through $\mathbf{O_3}$ with $\mathbf{n_{\beta t}} = \mathbf{R_3}\cdot\mathbf{C_3}^{-\mathbf{T}}\cdot\mathbf{e_t}$ and $d_{\beta t} = -v_{31}^{(t)}$.

Besides their geometrical properties (discussed earlier in section 2.3) these two groups of homography-matrices are of interest because of two things: *Firstly*, with the equations of section 2.3 it is possible to compute the missing epipoles:

$$\mathbf{v_{12}} = \mathbf{K_t^{-1}}\cdot\mathbf{v_{21}} \quad \Rightarrow \quad \mathbf{v_{32}} = \mathbf{J_t}\cdot\mathbf{v_{12}} \quad \textit{and} \quad \mathbf{v_{13}} = \mathbf{J_t^{-1}}\cdot\mathbf{v_{31}} \quad \Rightarrow \quad \mathbf{v_{23}} = \mathbf{K_t}\cdot\mathbf{v_{13}} \tag{4.7}$$

Because of the inversion in these computations at least one regular $\mathbf{J_t}$- and $\mathbf{K_t}$-matrix is needed. Since the determination whether a matrix is singular or not using noisy data is an ill-posed problem, a good way to overcome this is to choose that $\mathbf{J_t}$- and $\mathbf{K_t}$-matrix which has the best conditioning number because it is very unlikely (though possible) that all three $\mathbf{J_t}$- or $\mathbf{K_t}$-matrices are singular.

Secondly, using these $\mathbf{J_t}$- and $\mathbf{K_t}$-matrices and the computed epipoles we can determine the fundamental matrices between the views ψ_1 and ψ_2 and between ψ_1 and ψ_3 with equation (2.6). With these so derived fundamental matrices and the given IOR-matrices $\mathbf{C_1}$, $\mathbf{C_2}$, $\mathbf{C_3}$ we are able to compute the rotation matrices $\mathbf{R_2}$ and $\mathbf{R_3}$ (by setting $\mathbf{R_1} = \mathbf{I}$, i.e. by working with the relative orientation of successive images); [Brandstätter 1991]. In case of unknown IOR-matrices it is possible to compute a common IOR for all images ($\mathbf{C_1} = \mathbf{C_2} = \mathbf{C_3}$); [Niini 1994]. But for this purpose a fundamental matrix between ψ_2 and ψ_3 is necessary, which can be computed using a homography-matrix $\mathbf{H}$ between these images. A rank-2 homography $\mathbf{H_{23}}$ resp. $\mathbf{H_{32}}$ from image ψ_2 to image ψ_3 resp. from image ψ_3 to image ψ_2 can be computed by:

$$\mathbf{H_{23}} = \mathbf{J_t}\cdot\mathbf{K_s^{-1}} - \frac{1}{v_{31}^{(s)}}\cdot\mathbf{v_{31}}\cdot\mathbf{e_t^T} \quad \textit{resp.} \quad \mathbf{H_{32}} = \mathbf{K_t}\cdot\mathbf{J_s^{-1}} - \frac{1}{v_{21}^{(s)}}\cdot\mathbf{v_{21}}\cdot\mathbf{e_t^T} \quad \textit{with } s,\, t \in \{1,2,3\} \tag{4.8}$$

Whether $\mathbf{H_{23}}$ or $\mathbf{H_{32}}$ is to be chosen depends on which homography $\mathbf{J_s}$ or $\mathbf{K_s}$ is the best conditioned one. An IOR- and XOR-computation for a given tensor can be found in [Ressl 1997].

6 FUTURE WORK

Within this paper an introduction to the relative orientation of three images in a linear way by means of the trifocal tensor was given. This trifocal tensor is of interest because of three reasons: Firstly, the theoretical and geometrical background of the underlying relations (some of them have been reported in this paper); secondly, as a means for computing approximate values for the XOR and the common IOR of three images (i.e. image orientation and calibration); and thirdly, the trifocal tensor constitutes a first suitable means to detect blunders 'on a small scale' in advance, on the contrary to gross-error-detection in the whole image set by means of data-snooping.

In [Torr, Zisserman 1997] a method called RANSAC (random sample consensus) is used for the blunder-detection. In the course of that out of a large set of *n* point correspondences (order 100) a subset of six homologous points is selected and the trifocal tensor underlying this subset is computed uniquely. Afterwards the number of outliers is computed; i.e. those of the remaining *(n-6)* point-correspondences not being involved in the tensor's computation having an error above a certain threshold. After that the whole procedure is repeated for several (~500) other subsets of six points out of the given point correspondences. This is done to make sure (~95% probability) that there is at least one subset containing only good data points. Out of the resulting multitude of trifocal tensors the one having least outliers is chosen as an approximation for the following least-squares-adjustment including all the correspondences considered as inliers yielding an improved estimation.

Of further interest is the topic of dangerous surfaces. In [Shashua, Maybank 1996] it is shown, that if the trifocal tensor is computed by means of >6 homologous point-triples no such surfaces exist, but 10 certain singular positions for the points in space (including the projection centers). The arrangement of these dangerous points is not given, except that these points arise as the intersection (base points) of a linear system of cubic surfaces. Also not given is their 'sphere of influence', i.e. how far away points must lie to allow a solution. In this connection it must be mentioned that the

dimension of 'dangerous surfaces' in practice always appears higher by one. [Luong, Faugeras 1996] use in this context the term '*critical volume*' for emphasizing that depending on the image noise ambiguities and inaccuracies of the unknowns may also appear for configurations quite far away from the 'exact' dangerous surface. In [Maybank, Shashua 1998] it is shown that reconstruction from three images of six points is subject to a three way ambiguity which is preserved as long as the optical center of the camera remains on a certain quadric surface; i.e. if 6 points are used for computing the trifocal tensor a real dangerous surface can exist, which is of a special interest for the 6 point algorithm proposed by [Torr, Zisserman 1997]. The question arises if this ambiguity due to six points can be broken by any seventh point, or are there some constraints which such a 'point of deliverance' must meet.

The topics of gross-error-detection, dangerous situations, constraints within the TFT (also for calibrated images) and the inclusion of image-distortion into the TFT-framework will be investigated in more detail in the near future in the course of a research-project subsidized by the Austrian Science Fund FWF (P13901-INF).

REFERENCES

G. Brandstätter,1991. Notizen zur voraussetzungslosen gegenseitigen Orientierung von Meßbildern. Österreichische Zeitschrift für Vermessungswesen und Photogrammetrie, 79. Jahrgang, Heft 4, pp. 273-280.

R. Hartley, 1994. Lines and Points in Three Views – a Unified Approach. Proc. of an Image Understanding Workshop held in Monterey, California, Nov. 13-16, Vol. II, pp. 1009-1016.

R. Hartley, 1998. Computation of the Quadrifocal Tensor. Proc. of 5th European Conf. on Computer Vision, Springer.

K.R. Koch 1987. Parameterschätzung und Hypothesentests. 2. Aufl., Dümmler Bonn, pp. 174-188, 245.

H.C. Longuet-Higgins, 1981. A Computer Algorithm for Reconstructing a Scene from two Projections. Nature Vol. 293.

Q.-T. Luong, O.D. Faugeras, 1996. The Fundamental Matrix: Theory, Algorithms and Stability Analysis. The International Journal of Computer Vision, 1(17), pp. 43-76.

S. Maybank, A. Shashua, 1998. Ambiguity in Reconstruction From Images of Six Points. Proc. Int. Conf. on Computer Vision, IEEE Computer Society Press, pp. 703-708.

I. Niini, 1994. Relative Orientation of Multiple Images Using Projective Correlation. Helsinki University of Technology, Institute of Photogrammetry and Remote Sensing.

T. Papadopoulo, O.D. Faugeras, 1998. A New Characterization of the Trifocal Tensor. Proc. of 5th European Conf. on Computer Vision, edited by H. Burkhardt and B. Neumann, Springer.

C. Ressl, 1997. Relative Orientierung dreier Bilder mit Hilfe des trilinearen Tensors. Diploma thesis; Institute of Photogrammetry and Remote Sensing, University of Technology, Vienna.

K. Rinner, 1963. Studien über eine allgemeine, voraussetzungslose Lösung des Folgebildanschlusses. Sonderheft 23 der Österreichischen Zeitschrift für Vermessugswesen.

K. Rinner, R. Burkhardt, 1972. Photogrammetrie. In: Handbuch der Vermessungskunde. (Hsgb. Jordan, Eggert, Kneissel) Band III a/3. Stuttgart: J.B. Metzlersche Verlagsbuchhandlung, pp. 2286-.

A. Shashua, 1995. Algebraic Functions For Recognition. PAMI Vol. 17 No. 8, pp. 779-788.

A. Shashua, S. Maybank, 1996. Degenerate N Point Configurations of Three Views: Do Critical Surfaces Exist? Technical Report TR 96-19, Hebrew University of Jerusalem.

A. Shashua, M. Werman, 1995. On the Trilinear Tensor of Three Perspective Views and its Underlying Geometry. Proc. of the International Conference on Computer Vision (ICCV), Boston MA.

M.E. Spetsakis, J. Aloimonos, 1990. A Unified Theory of Structure from Motion. Proc. of an Image Understanding Workshop held at Pittsburgh, Pennsylvania, Sept. 11-13, pp. 271-283.