Journal for Geometry and Graphics
Volume 10 (2006), No. 2, 137–153.

Descriptive Geometry Meets Computer Vision – The Geometry of Two Images

Hellmuth Stachel

Institute of Discrete Mathematics and Geometry, Vienna University of Technology
Wiedner Hauptstr. 8-10/104, A 1230 Vienna, Austria
email: stachel@dmg.tuwien.ac.at

Abstract. The geometry of multiple images has been a standard topic in Descriptive Geometry and Photogrammetry (Remote Sensing) for more than 100 years. During the last twenty years great progress has been made within the field of Computer Vision, a topic with the main goal to endow a computer with a sense of vision. The previously graphical or mechanical methods of reconstruction have been replaced by mathematical methods as offered by computer algebra systems. This paper will explain to geometers how to reconstruct two digital images of the same scene and how to recover metrical data of the depicted object — using standard software only. Not the presented results are new, but the way how they are deduced by geometric reasoning. The arguments are based on Linear Algebra and classical Descriptive Geometry results.

Key Words: Descriptive Geometry, multiple images, two-views-system, essential matrix, Computer Vision

MSC: 51N05

1. Introduction

1.1. Central perspectives

The basic term in this paper is the *central projection* with *center* $\mathbf{c}$ and *image plane* π (see Fig. 1). This is the geometric idealization of the photographic mapping with $\mathbf{c}$ as the focal point or focal center of the lenses and π as the plane carrying the film or the CCD sensor. The pedal point of $\mathbf{c}$ with respect to π is called *principal point* $\mathbf{h}$; the distance $d := \|\mathbf{c} - \mathbf{h}\|$ is the *focal length.* The image is called *(central* or *linear) perspective.*

Each central projection or photographic mapping defines a particular coordinate system in space, the *camera frame.* Its origin is placed at the center $\mathbf{c}$, the principal ray of the camera is the $\overline{z}$-axis. And the principal directions in the photosensitive plane serve as $\overline{x}$- and $\overline{y}$-axis. These coordinate axes span the vanishing plane π_v of this central projection.

ISSN 1433-8157/\$ 2.50

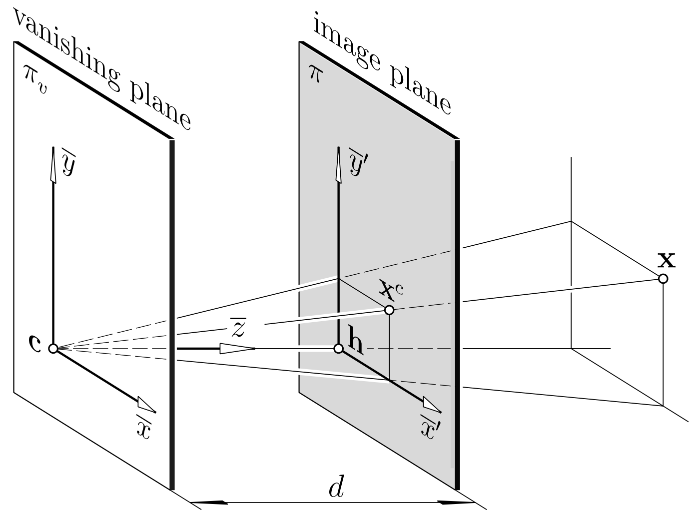

Figure 1: Central projection with center $\mathbf{c}$, principal point $\mathbf{h}$ and camera frame $\overline{x}, \overline{y}, \overline{z}$

When at the same time the principal point $\mathbf{h}$ is the origin of 2D-coordinates $(\overline{x}', \overline{y}')$ in the image plane, then the photographic mapping $\mathbf{x} \mapsto \mathbf{x}'$ obeys the matrix equation

$$\begin{pmatrix} \overline{x}' \\ \overline{y}' \end{pmatrix} = \frac{d}{\overline{z}} \begin{pmatrix} \overline{x} \\ \overline{y} \end{pmatrix}.$$

It is appropriate to introduce homogeneous 2D-coordinates $(\overline{x}'_0 : \overline{x}'_1 : \overline{x}'_2)$ by

$$\overline{x}' = \frac{\overline{x}'_1}{\overline{x}'_0}, \quad \overline{y}' = \frac{\overline{x}'_2}{\overline{x}'_0}.$$

In the same way we use homogeneous 3D-coordinates obeying

$$(\overline{x}_0 : \overline{x}_1 : \overline{x}_2 : \overline{x}_3) = (1 : \overline{x} : \overline{y} : \overline{z}).$$

Then the central projection is expressed as a linear mapping[1]

$$\begin{pmatrix} \overline{x}'_0 \\ \overline{x}'_1 \\ \overline{x}'_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & d & 0 & 0 \\ 0 & 0 & d & 0 \end{pmatrix} \cdot \begin{pmatrix} \overline{x}_0 \\ \vdots \\ \overline{x}_3 \end{pmatrix}.$$

Now we bring this in a more general form: We replace the camera frame by arbitrary *world coordinates* (x, y, z). And we admit that in the image plane π our particular frame is modified by a translation and by scalings to the system of (x', y')-coordinates. In this way we obtain the general form of mapping equations for central projections:

$$\begin{pmatrix} x'_0 \\ x'_1 \\ x'_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ h'_x & df_x & 0 \\ h'_y & 0 & df_y \end{pmatrix}}_{\substack{\text{intrinsic calibration} \\ \text{parameters}}} \cdot \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdot \underbrace{\begin{pmatrix} 1 & 0 \quad 0 \quad 0 \\ z_x & \\ z_y & R \\ z_z & \end{pmatrix}}_{\substack{\text{extrinsic calibration} \\ \text{parameters}}} \cdot \begin{pmatrix} x_0 \\ \vdots \\ x_3 \end{pmatrix}. \tag{1}$$

[1]Throughout this paper coordinate vectors are seen as column vectors.

On the right hand side there is a triple product of matrices. The first matrix contains beside the focal distance d the new image coordinates (h'_x, h'_y) of the principal point $\mathbf{h}$ and the two scaling factors f_x, f_y which usually are set to 1. These entries are called the *intrinsic calibration parameters* of the photo. If these parameters are known the photo is called *calibrated*. In this case the image determines the bundle of rays $\mathbf{c} \vee \mathbf{x}$ up to a rigid spatial motion.

The last matrix in the triple product of eq. (1) contains the orthogonal (3×3)-matrix R and the world coordinates (z_x, z_y, z_z) of the center $\mathbf{c}$. This defines the position of the camera frame with respect to the world coordinates. The involved entries are called *extrinsic calibration parameters*.

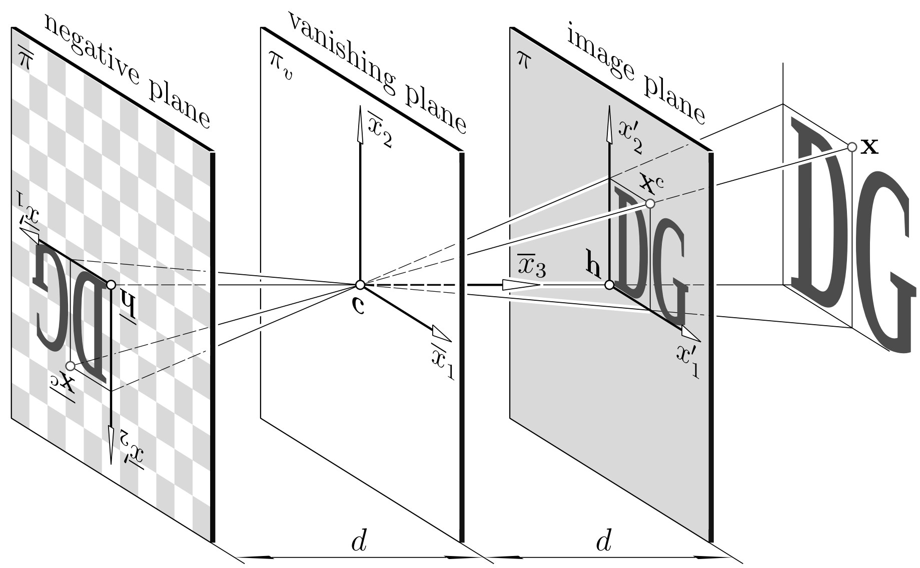

Figure 2: Central projection into the negative plane

Fig. 2 reveals that it does not matter, whether the image plane π is located between the center $\mathbf{c}$ and the scene (like in Fig. 1) or π is outside (like $\overline{\pi}$ in Fig. 2). The latter takes place at the photographic mapping. As long as the distance d is the same and the principal rays $\mathbf{c} \vee \mathbf{h}$ coincide, the images are congruent, provided, the image planes are seen from the correct side.

1.2. Linear images

We can generalize the central projection by a *central axonometry* (see, e.g., [13, 2, 12]). It maps the 3-space by a (singular) collinear transformation into the image plane. Hence, collinearity of points remains invariant and cross ratios are preserved. In homogeneous coordinates a central axonometry can again be expressed by a linear mapping; in matrix form like in (1) there is a (4×4)-matrix of rank 3. Therefore these images are called *linear images*.[2] There are several results on how to characterize central perspectives among linear images (see, e.g., [6, 13, 7, 11, 2, 12, 8]).

In the generic case linear images are *uncalibrated*. Such a linear image can, e.g., be obtained by taking a photo of a given photo. It can be proved (cf. [13]) that a linear image of a scene is always an *affine* transform of a central perspective of the same scene (compare the two views in Fig. 3).

[2] For a coordinate-free definition of linear images see [1].

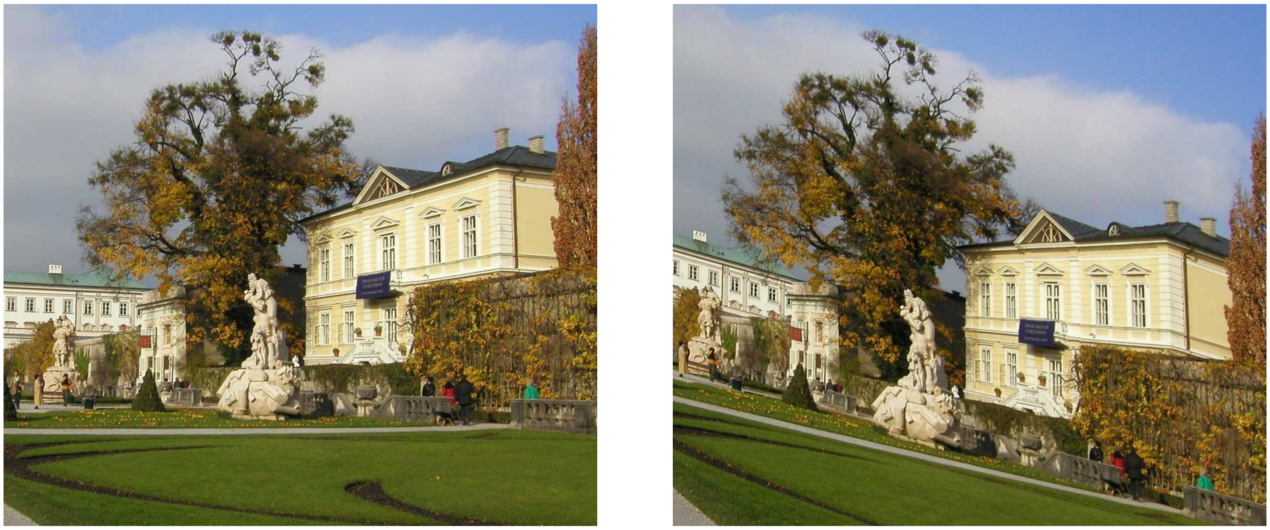

Figure 3: Central perspective (left) versus linear image (right)

According to our definition, a photo is uncalibrated as soon as the exact position of the center $\mathbf{c}$ of projection with respect to the photo is unknown. This is because replacing the exact center by any other point means that the bundle of rays $\mathbf{c} \vee \mathbf{x}$ connecting the center with points of the scene is replaced by a collinear transform of the original bundle.

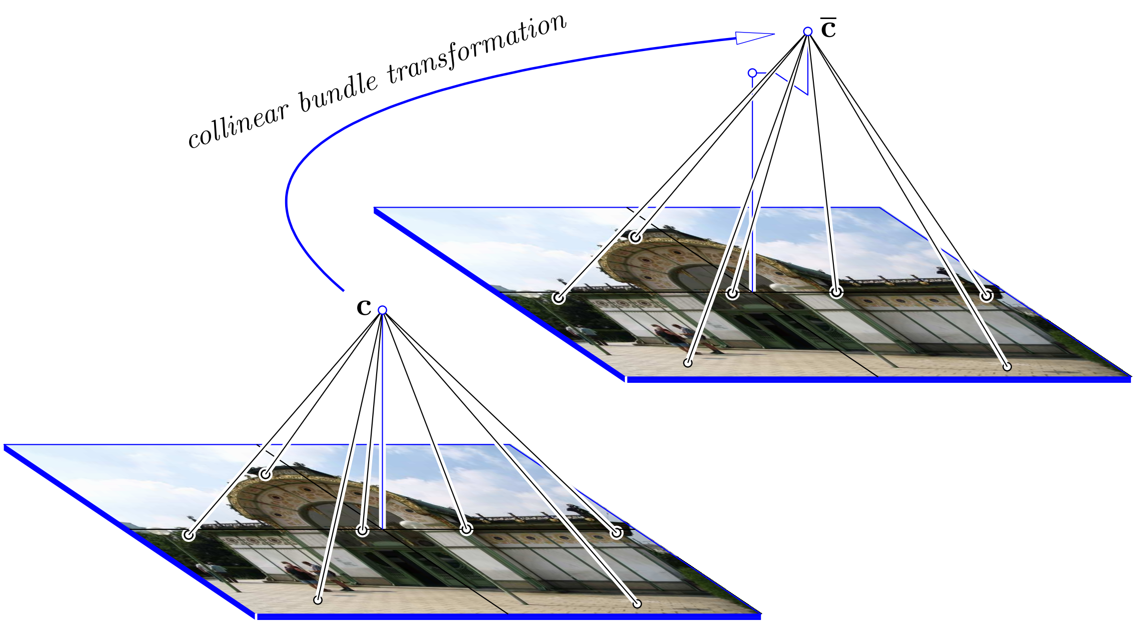

Figure 4: Replacing the original center $\mathbf{c}$ by $\overline{\mathbf{c}}$ acts like a collinear transformation on the bundle of rays

1.3. Singular value decomposition

One technical tool from Linear Algebra, which will be used in the sequel, is the *singular value decomposition* of any matrix. It decomposes the $(m\times n)$-matrix A into the product

$$A = U \cdot \mathrm{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0) \cdot V^T \quad \text{with} \quad \lambda_1, \ldots, \lambda_r > 0, \quad r \leq \min\{m, n\},$$

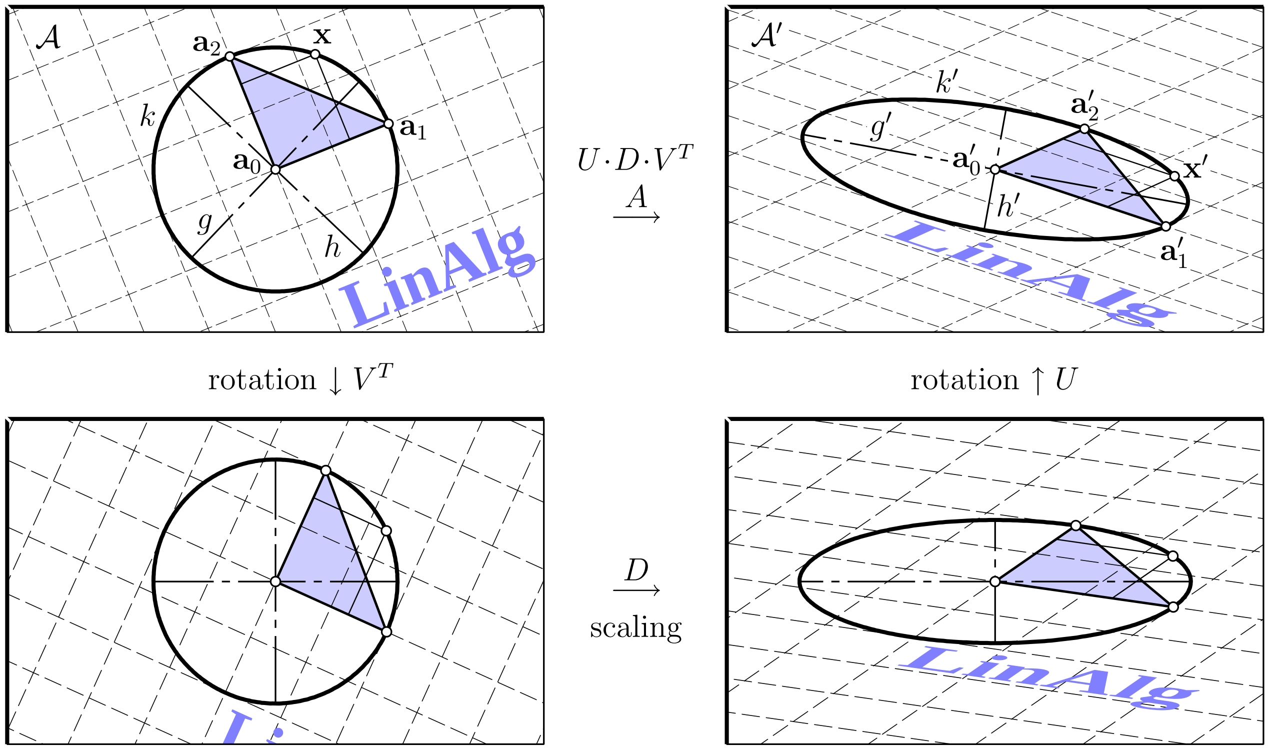

Figure 5: Singular value decomposition

of a diagonal $(m{\times}n)$-matrix and two orthogonal matrices U, V, i.e., $U^{-1} = U^T$ and $V^{-1} = V^T$. The non-zero entries $\lambda_1, \dots, \lambda_r$ in the main diagonal of the central factor are called the *singular values* of A. They are the positive square roots of the non-zero eigenvalues of the symmetric matrix $A^T{\cdot}A$ and therefore uniquely determined.

There is an instructive geometric interpretation of this decomposition in dimension 2 which can easily be generalized to the Euclidean n-space: Matrix A represents an affine transformation $\mathcal{A} \to \mathcal{A}'$ (see Fig. 5) mapping the unit circle k onto an ellipse k', which might also be degenerated. There are pairwise orthogonal diameters g, h of the unit circle which are mapped onto the axes of symmetry g', h' of the corresponding ellipse. These particular pairs of diameters define the directions of *principal distortions* for this affine map.

The singular values of A equal the semiaxes of the ellipse. Therefore the singular values are sometimes called the *principal distortion ratios* of this affine map. The orthogonal matrices U and V^T represent the coordinate transformations between the given frames and that of the principal distortion directions.

2. The geometry of two images

The geometry of pairs of central views has been a classical topic in Descriptive Geometry. Important results are, e.g., due to S. Finsterwalder, E. Kruppa [9], J. Krames, W. Wunderlich, H. Brauner [1].

2.1. Uncalibrated case

Let two central projections be given with centers $\mathbf{c}_i$ and image planes π_i, $i = 1, 2$. This refers to the *viewing situation* in 3-space as displayed in Fig. 6. In addition, let κ_1, κ_2 be

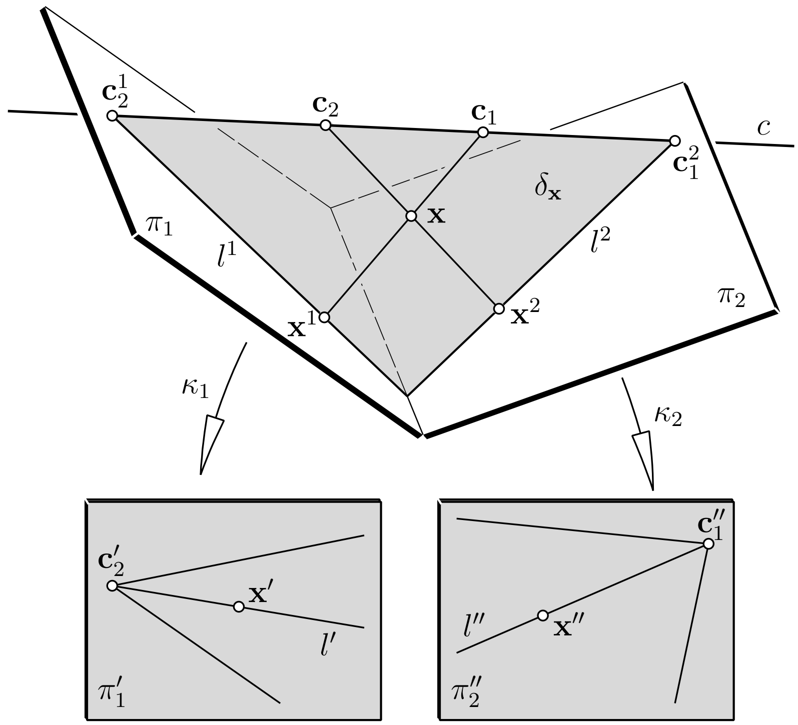

Figure 6: Epipolar constraint in a two-views-system

collinear transformations which map the images into π_1' and π_2'', respectively. In this way we have defined a *general two-views-system* consisting of two linear images. Any space point $\mathbf{x}$ different from the two centers is represented by its two views $\mathbf{x}', \mathbf{x}''$. We call these two images of $\mathbf{x}$ *corresponding*.

The basic geometric property of two-views-systems results from the fact that for space points $\mathbf{x}$ which are not aligned with the two centers, the two rays of sight $\mathbf{c}_1 \vee \mathbf{x}$ and $\mathbf{c}_2 \vee \mathbf{x}$ are coplanar (see Fig. 6). They are located in a plane $\delta_{\mathbf{x}}$ which in both linear images appears in an edge view. In the viewing situation the images of the pencil of planes $\delta_{\mathbf{x}}$ constitute two *perspective* line pencils. After applying the collinear transformations κ_1, κ_2 there remain *projective* pencils of lines, the so-called *epipolar lines*. The centers $\mathbf{c}_2'$ and $\mathbf{c}_1''$ of these pencils are called *epipoles*. As expressed in the notation, each epipole is the image of one center under the other projection. The projectivity between the two pencils is called *epipolar constraint*. We summarize:

Theorem 1 1. *For any two linear images of a scene there is a projectivity between two particular line pencils*

$$\mathbf{c}_2'(\delta_{\mathbf{x}}') \barwedge \mathbf{c}_1''(\delta_{\mathbf{x}}'')$$

such that two points $\mathbf{x}', \mathbf{x}''$ are corresponding, i.e., images of the same space point, if and only if they are located on corresponding epipolar lines.

2. *Using homogeneous coordinates, there is a matrix $B = (b_{ij})$ of rank 2 such that two points $\mathbf{x}' = (x_0' : x_1' : x_2')$ and $\mathbf{x}'' = (x_0'' : x_1'' : x_2'')$ are corresponding if and only if*

$$\sum_{i,j=0}^{2} b_{ij}\, x_i'\, x_j'' = \mathbf{x}'^T \cdot B \cdot \mathbf{x}'' = 0\,. \tag{2}$$

Remark: This vanishing bilinear form defines a correlation which is singular because of the rank deficiency of the so-called *essential matrix* B.

Proof: Using homogeneous line coordinates, the projectivity between the line pencils can be expressed by

$$(\mathbf{u}'_1\lambda_1 + \mathbf{u}'_2\lambda_2)\mathbb{R} \;\mapsto\; (\mathbf{u}''_1\lambda_1 + \mathbf{u}''_2\lambda_2)\mathbb{R}$$

for $(\lambda_1,\lambda_2) \in \mathbb{R}^2 \setminus \{(0,0)\}$, provided $\mathbf{u}'_1,\dots,\mathbf{u}''_2$ are particularily normalized line coordinates of two pairs of corresponding epipolar lines.
The points $\mathbf{x}'$ and $\mathbf{x}''$ are corresponding, i.e., images of the same space point, iff there is a nontrivial pair (λ_1,λ_2) such that

$$\begin{aligned}(\mathbf{u}'_1\lambda_1 + \mathbf{u}'_2\lambda_2)\cdot \mathbf{x}' &= 0\\ (\mathbf{u}''_1\lambda_1 + \mathbf{u}''_2\lambda_2)\cdot \mathbf{x}'' &= 0\,.\end{aligned}$$

These two linear homogeneous equations in the unknowns (λ_1,λ_2) have a nontrivial solution if and only if the determinant vanishes. This gives the stated bilinear form

$$(\mathbf{u}'_1\cdot\mathbf{x}')(\mathbf{u}''_2\cdot\mathbf{x}'') - (\mathbf{u}'_2\cdot\mathbf{x}')(\mathbf{u}''_1\cdot\mathbf{x}'') = 0$$

or — in matrix form —

$$\mathbf{x}'^T\cdot(\mathbf{u}'_1\cdot\mathbf{u}''^T_2 - \mathbf{u}'_2\cdot\mathbf{u}''^T_1)\cdot\mathbf{x}'' = \mathbf{x}'^T\cdot B\cdot\mathbf{x}'' = 0\,.$$

There are singular points of this correspondance: The epipole $\mathbf{c}'_2$ corresponds to all $\mathbf{x}''$, i.e., $\mathbf{c}'^T_2\cdot B\cdot\mathbf{x}'' = 0$ for all $\mathbf{x}'' \in \mathbb{R}^3$, therefore $\mathbf{c}'^T_2\cdot B = \mathbf{0}$. Vice versa, all points $\mathbf{x}' \in \mathbb{R}^3$ correspond to $\mathbf{c}''_1$, i.e., $B\cdot\mathbf{c}''_1 = \mathbf{0}$. As these homogeneous linear systems have a one-dimensional solution, the essential (3×3)-matrix B has rank 2. □

2.2. Calibrated case

In the calibrated case we can express the essential matrix B in a particular form. For this purpose it is necessary to specify the homogeneous coordinates used in the bilinear relation (2): For each image point we take its 3D coordinates with respect to the camera frame as homogeneous 2D coordinates (see Fig. 7).

Theorem 2 *When in the calibrated case the camera frame coordinates serve as homogeneous coordinates of the image points* $\mathbf{x}', \mathbf{x}''$*, then the essential matrix* B *is the product of a skew-symmetric matrix and an orthogonal one, i.e.,*

$$B = S\cdot R \text{ with } S^T = -S \text{ and } R^{-1} = R^T. \tag{3}$$

Then the two singular values of B *are equal.*

Proof: According to Fig. 7, the three vectors

$$\mathbf{c}' := \mathbf{c}_2 - \mathbf{c}_1\,,\quad \mathbf{x}' \quad\text{and}\quad \mathbf{x}''$$

are coplanar. Therefore their triple product vanishes. However, we have to pay attention to the fact that $\mathbf{x}'$ and $\mathbf{x}''$ are given in two different camera frames. Let

$$\mathbf{x}_1 = \mathbf{c}' + R\cdot\mathbf{x}_2 \tag{4}$$

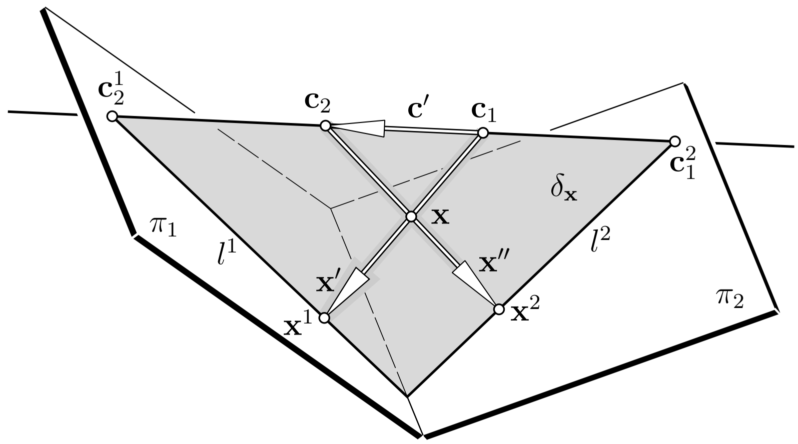

Figure 7: Epipolar constraints for calibrated images

be the coordinate transformation from the second camera frame into the first one with an orthogonal matrix R. Now the complanarity is equivalent to

$$0 = \det(\mathbf{x}',\ \mathbf{c}',\ R\cdot\mathbf{x}'') = \mathbf{x}' \cdot (\mathbf{c}' \times R\cdot\mathbf{x}'').$$

We may replace the cross product by the product of $\mathbf{x}''$ with a skew-symmetric matrix, i.e.,

$$\mathbf{c}' \times R\cdot\mathbf{x}'' = S \cdot R \cdot \mathbf{x}''$$

and

$$S = \begin{pmatrix} 0 & -z'_z & z'_y \\ z'_z & 0 & -z'_x \\ -z'_y & z'_x & 0 \end{pmatrix}, \tag{5}$$

provided (z'_x, z'_y, z'_z) are the coordinates of $\mathbf{c}_2$ with respect to the first camera frame. It is noteworthy that according to (4) the two factors S and R together define the relative position between the two camera frames uniquely.

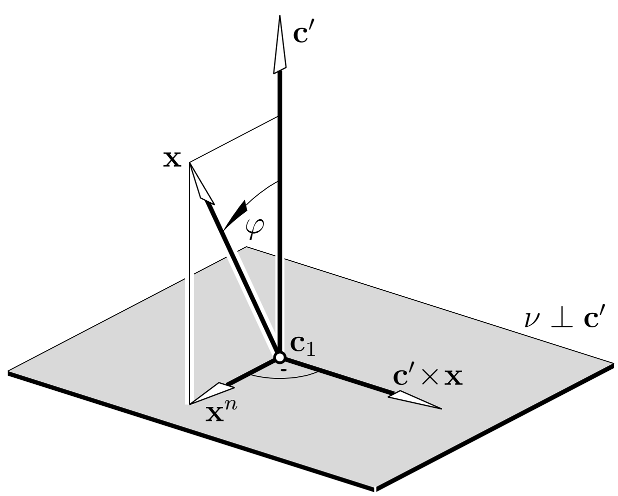

Figure 8: $\mathbf{x} \mapsto S\cdot\mathbf{x} = \mathbf{c}' \times \mathbf{x}$ is the product of an orthogonal projection, a 90°-rotation, and a scaling with factor $\|\mathbf{c}'\|$

The singular values of $B = S\cdot R$ can either be computed straight forward as the positive squareroots of eigenvalues of $B^T\cdot B$, i.e., of $S^T\cdot S = -S\cdot S$. But we can also proceed in a more geometric way by understanding B as matrix of an affine transformation in 3-space (see Fig. 8): The cross product $\mathbf{c}'\times\mathbf{x}$ is orthogonal to the plane spanned by $\mathbf{c}'$ and $\mathbf{x}$, and it has the length

$$\|\mathbf{c}'\times\mathbf{x}\| = \|\mathbf{c}'\|\,\|\mathbf{x}\|\sin\varphi = \|\mathbf{c}'\|\,\|\mathbf{x}^n\|$$

where $\mathbf{x}^n$ is the orthogonal projection of $\mathbf{x}$ in direction of $\mathbf{c}'$. So, the mapping $\mathbf{x}\mapsto S\cdot\mathbf{x}$ is the composition of an orthogonal projection, of a 90°-rotation, and a scaling with factor $\|\mathbf{c}'\|$. As the singular values (= principal distortion ratios) of an orthogonal projection are $(1,1)$ (note Fig. 5), the singular values of S are $(\|\mathbf{c}'\|, \|\mathbf{c}'\|)$. □

3. The fundamental theorems

What means *'reconstruction'* from two images? The photos have been taken in particular positions of the camera. These poses will be called *viewing situation*. But afterwards we have only the two images, and we know nothing about how the camera frames where mutually placed in 3-space. Hence, any reconstruction includes both, recovering the viewing situation and recovering the depicted scene.

The problem of recovering a scene from two or more images is a basic problem in Computer Vision (see, e.g., [3, 4, 5, 15]). It is remarkable, that sometimes in the cited books the authors actually acknowledge results which have already been achieved in Descriptive Geometry (note, e.g., the high estimation of E. Kruppa's results [9] in [15]). However, Computer Vision focuses on numerical solutions, and the use of computers brought new insight and progress in this problem. Since measuring pixels in any image can be carried out with standard software, it has become possible to recover an object with high precision from two digital images just by using a laptop.

Theorem 3 *From two uncalibrated images with given projectivity between epipolar lines the depicted object can be reconstructed up to a collinear transformation.*

Sketch of the *proof:* The two images can be placed in space such that pairs of epipolar lines are intersecting. For this purpose we start with a position where the two images are coplanar and two corresponding lines are aligned. Then the two pencils of epipolar lines are perspective with respect to an axis a. Now we rotate one of the image planes about this axis a. The corresponding epipolar lines are still intersecting on a. Then we specify arbitrary centers $\mathbf{c}_1$, $\mathbf{c}_2$ on the baseline c which connects the two epipoles. This gives rise to a reconstructed 3D object.

Now it remains to prove in detail that any other choice of a viewing situation — with intersecting pairs of epipolar lines but different centers — results in a recovered object which is a collinear transform of the previous one. □

Theorem 4 [S. Finsterwalder, 1899] *From two calibrated images with given projectivity of epipolar lines the depicted object can be reconstructed up to a similarity.*

Sketch of the *proof:* For the two projections the pencils of epipolar planes $\delta_{\mathbf{x}}$ need to be *congruent*. There is a rigid motion from one camera frame to the other such that any two corresponding epipolar planes become coincident. Then for any choice of $\mathbf{c}_2$ relative to $\mathbf{c}_1$ on the carrier line c of the unified pencil of planes there exists a reconstructed 3D object. Now it is obvious that any other choice of $\mathbf{c}_2$ on line c gives a similar 3D object. □

In this sense the problem of recovering a scene is reduced to the determination of epipoles. This problem is equivalent to a classical problem of Projective Geometry, the *'Problem of Projectivity'* (see Fig. 9):

Given: 7 pairs of corresponding points $(\mathbf{x}'_1, \mathbf{x}''_1), \dots, (\mathbf{x}'_7, \mathbf{x}''_7)$.

Wanted: A pair of points $(\mathbf{s}', \mathbf{s}'')$ (= epipoles) such that for $i = 1, \dots, 7$ the connecting lines $\mathbf{s}' \vee \mathbf{x}'_i$ and $\mathbf{s}'' \vee \mathbf{x}''_i$ are corresponding under a projectivity, i.e.,

$$\mathbf{s}'(\mathbf{s}' \vee \mathbf{x}'_i) \;\barwedge\; \mathbf{s}''(\mathbf{s}'' \vee \mathbf{x}''_i), \quad i = 1, \dots, 7\,.$$

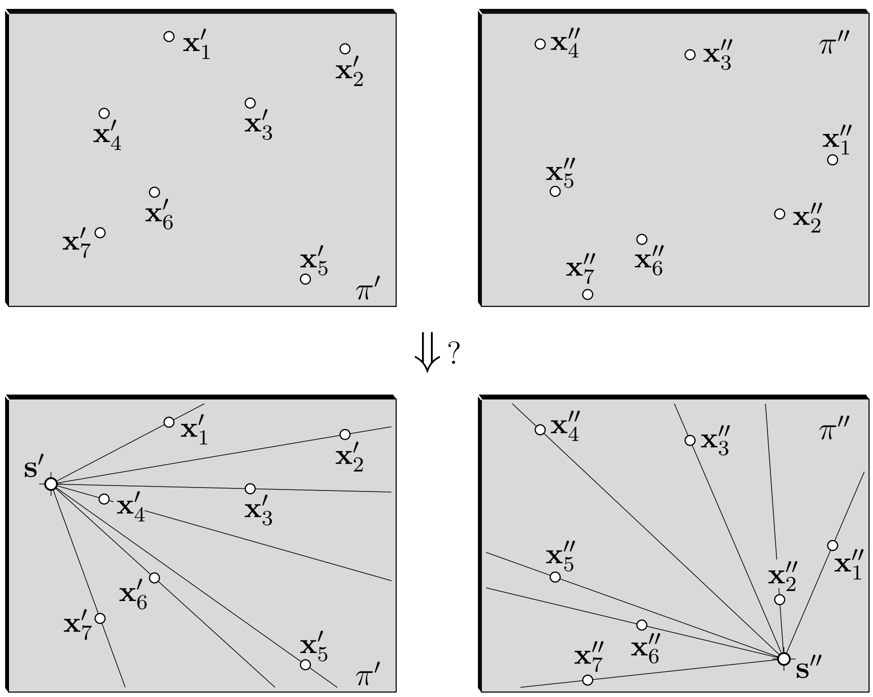

Figure 9: The Problem of Projectivity

The Problem of Projectivity is a *cubic* problem. This follows from the following reasoning: Due to eq. (2) the 7 given pairs of corresponding points give $n = 7$ linear homogeneous equations

$$\mathbf{x}'^T_i \cdot B \cdot \mathbf{x}''_i = 0, \quad i = 1, \dots, n\,, \tag{6}$$

for the 9 entries in the essential (3×3)-matrix $B = (b_{ij})$. The condition $\mathrm{rk}(B) = 2$ gives the additional cubic equation $\det B = 0$ which fixes all b_{ij} up to a common factor.

4. Computing the best fitting essential matrix

For noisy image points it is recommended to use $n > 7$ pairs of corresponding points, so-called *reference points* (see Fig. 11), and to apply methods of least squares approximation for obtaining the 'best fitting' essential matrix B. This is done in two steps:

Step 1: Let A denote the $(n\times 9)$ coefficient matrix in the linear system (6) of homogeneous equations for the entries of B. Then the 'least square fit' $\widetilde{B}$, i.e., the 'best' solution, is an

Figure 10: Given photos: Historical 'Stadtbahn' station Karlsplatz in Vienna (Otto Wagner, 1897)

Figure 11: Identifying 20 reference points

eigenvector to the smallest eigenvalue of the symmetric matrix $A^T \cdot A$ which minimizes the quadratic form

$$\mathbf{y}^T \cdot A^T \cdot A \cdot \mathbf{y} = \|A \cdot \mathbf{y}\|^2$$

under the side condition $\|\mathbf{y}\| = 1$.[3]

Step 2: Any essential matrix has rank 2, and in particular for the calibrated case the two singular values must be equal. In order to obtain such a 'best fitting' essential matrix B for our obtained $\widetilde{B}$, we use what sometimes is called the *'projection into the essential space':* This is based on the singular value decomposition of $\widetilde{B}$, which has been presented in Section 1. We factorize $\widetilde{B}$ into a matrix product

$$\widetilde{B} = U \cdot D \cdot V^T, \quad D = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3) \text{ with } \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0 \tag{7}$$

and with orthogonal U, V.

[3]In the sense of Fig. 5 the requested eigenvector spans a diameter of the unit sphere (k) which is mapped onto the diameter carrying the shortest semi-axis of the corresponding ellipsoid (k'). This diameter is orthogonal to the best fitting hyperplane.

Theorem 5 *Let $\widetilde{B} = U \cdot \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3) \cdot V^T$ by (7) be the singular value decomposition of the best solution $\widetilde{B}$ obeying the homogeneous linear system of equations (6). Then*

1. *in the uncalibrated case the matrix*

$$B = U \cdot \mathrm{diag}(\lambda_1, \lambda_2, 0) \cdot V^T \tag{8}$$

is best in the sense of the Frobenius norm, i.e., $\|\widetilde{B} - B\|_f = \lambda_3$ is minimal, provided $\lambda_3 < \lambda_2 \le \lambda_1$. Otherwise the minimum is not unique. For the sake of completeness we add:

2. *If λ_3 is a twofold singular value of $\widetilde{B}$, then all $B = U \cdot A \cdot V^T$ with*

$$A = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_3(1-n_2^2) & -\lambda_3 n_2 n_3 \\ 0 & -\lambda_3 n_2 n_3 & \lambda_3(1-n_3^2) \end{pmatrix} \quad \text{and} \quad n_2^2 + n_3^2 = 1$$

reach the minimal norm $\|\widetilde{B} - B\|_f = \lambda_3$. If the singular value λ_3 has multiplicity 3 then even all

$$A = \lambda_3(I_3 - \mathbf{n}\cdot\mathbf{n}^T), \ \mathbf{n} \in \mathbb{R}^3 \ \text{with} \ \|\mathbf{n}\| = 1 \ \text{and the unit matrix} \ I_3\,,$$

are minimal. Here A represents the product of an orthogonal projection in direction of $\mathbf{n}$ and a scaling.

3. *In the calibrated case with the particular homogeneous coordinates according to Theorem 2 the best matrix is*

$$B = U \cdot \mathrm{diag}(\lambda, \lambda, 0) \cdot V^T \ \text{with} \ \lambda = \frac{\lambda_1 + \lambda_2}{2}\,. \tag{9}$$

Proof: For the square matrix $A = (a_{ij})$ the Frobenius norm $\|A\|_f = \sqrt{\sum_{i,j} a_{ij}^2}$ equals the trace of $A^T \cdot A$ and therefore the square sum of the singular values of A (see, e.g., [10, 15]). In the uncalibrated case the product B by (8) gives

$$\|\widetilde{B} - B\|_f = \|\, U \cdot [\,\mathrm{diag}(\lambda_1, \lambda_2, \lambda_3) - \mathrm{diag}(\lambda_1, \lambda_2, 0)] \cdot V^T \|_f = \|\mathrm{diag}(0, 0, \lambda_3)\|_f = \lambda_3\,.$$

In order to figure out which $B = U \cdot A \cdot V^T$ gives the minimal norm $\|\widetilde{B} - B\|_f$ among all rank 2 matrices, we use LAGRANGE's method for obtaining the minimum of the objective function

$$\zeta(A) := \|\widetilde{B} - B\|_f^2 = \sum_{j=1}^{3} (a_{jj} - \lambda_j)^2 + \sum_{i \ne j} a_{ij}^2$$

under the side condition $\det A = 0$. According to this the ten equations

$$\frac{\partial}{\partial a_{ij}} [\zeta(A) + \lambda \det A] = 0, \ \ i, j = 1, 2, 3, \quad \text{and} \ \ \det A = 0$$

give necessary conditions for the nine entries of A and the multiplier λ.

We rewrite these equations in vector form. For this purpose let $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ denote the standard basis and let $\mathbf{a}_j$, $j = 1, 2, 3$, be the column vectors of A. Furthermore we set

$$\mathbf{n}_k = \mathbf{a}_i \times \mathbf{a}_j \ \ \text{for} \ \ (i, j, k) = (1, 2, 3), \ (2, 3, 1), \ (3, 1, 2).$$

Now $\det A = 0$ is equivalent to the linear dependence of $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$. Hence there is a unit vector $\mathbf{n}$ such that

$$\mathbf{n}_k = \mathbf{a}_i \times \mathbf{a}_j = 2\mu_k \mathbf{n}, \quad \|\mathbf{n}\| = 1\,.$$

On the other hand we have $\det A = \mathbf{a}_j \cdot \mathbf{n}_j$ for each j, and the vanishing partial derivatives of our objective function $\zeta(A)$ are

$$\begin{array}{rl} \text{for diagonal entries} & 2(a_{jj} - \lambda_j) + \lambda(\mathbf{n}_j \cdot \mathbf{e}_j) = 0 \\ \text{and otherwise } (i \neq j) & 2a_{ij} + \lambda(\mathbf{n}_j \cdot \mathbf{e}_i) = 0\,. \end{array}$$

They can be combined in

$$2(\mathbf{a}_j - \lambda_j \mathbf{e}_j) = -\lambda \mathbf{n}_j = -2\lambda\mu_j \mathbf{n}\,.$$

For the Lagrange multiplier we conclude $\lambda \neq 0$ as otherwise the column vectors $\mathbf{a}_j = \lambda_j \mathbf{e}_j$ would be linearly independent. So, our ten equations are solved by

$$\mathbf{a}_j = \lambda_j \mathbf{e}_j - \lambda\mu_j \mathbf{n}\,, \tag{10}$$

provided the unknowns $\mu_1, \mu_2, \mu_3, \lambda$ and the coordinates of the unit vector $\mathbf{n}$ obey $\mathbf{n}_k = \mathbf{a}_i \times \mathbf{a}_j$ for all even permutations (i, j, k). We substitute (10) and obtain

$$2\mu_k \mathbf{n} = \lambda\lambda_i\mu_j(\mathbf{n} \times \mathbf{e}_i) - \lambda\lambda_j\mu_i(\mathbf{n} \times \mathbf{e}_j) + \lambda_i\lambda_j \mathbf{e}_k\,. \tag{11}$$

Expressed in coordinates (n_1, n_2, n_3) of $\mathbf{n}$ obeying $n_1^2 + n_2^2 + n_3^2 = 1$ we get

$$2\mu_k n_i = \lambda\lambda_j\mu_i n_k \quad \text{and} \quad 2\mu_k n_j = \lambda\lambda_i\mu_j n_k\,, \ \text{ i.e.,} \quad 2\mu_i n_k = \lambda\lambda_j\mu_k n_i\,, \tag{12}$$

and

$$\lambda\lambda_j\mu_i n_i + \lambda\lambda_i\mu_j n_j + 2\mu_k n_k = 2\lambda_i\lambda_j\,. \tag{13}$$

In a lengthy case analysis we have to distinguish whether for any j the pairs of homogeneous equations (12) for $\mu_i n_k$ and $\mu_k n_i$ are linearly dependent, i.e., $\lambda^2\lambda_j^2 = 4$, or there is only the trivial solution

$$\mu_1 n_2 = \mu_2 n_1 = \mu_2 n_3 = \mu_3 n_2 = \mu_3 n_1 = \mu_1 n_3 = 0.$$

We skip the first case which leads to infinite solutions as listed in 2.; this first case is only possible when the multiplicity of the smallest singular value is > 1. In the latter case we may first assume $n_3 \neq 0$ which implies

$$\mu_1 = \mu_2 = 0, \quad \mu_3 \neq 0, \quad \mathbf{n} = \mathbf{e}_3\,, \quad \mu_3 = \frac{\lambda_1\lambda_2}{2}\,, \quad \lambda = \frac{2\lambda_3}{\lambda_1\lambda_2}\,.$$

From (10) we obtain

$$\mathbf{a}_1 = \lambda_1 \mathbf{e}_1\,, \quad \mathbf{a}_2 = \lambda_2 \mathbf{e}_2\,, \quad \mathbf{a}_3 = \lambda_3 \mathbf{e}_3 - \lambda_3 \mathbf{e}_3 = \mathbf{0}.$$

This gives by (10) the minimal

$$\|\widetilde{B} - B\|_f^2 = \sum_{j=1}^{3} (\mathbf{a}_j - \lambda_j \mathbf{e}_j)^2 = \lambda^2 \sum_{j=1}^{3} \mu_j^2 = \lambda_3^2\,.$$

The alternative assumptions $n_1 \neq 0$ or $n_2 \neq 0$ result in larger values $\|\widetilde{B} - B\|_f = \lambda_1$ or λ_2.

Ad 3.: In the calibrated case we follow the proof given in [15, p. 119–120]: For any $B = U{\cdot}A{\cdot}V^T$ and $A = U' \cdot \mathrm{diag}(\lambda, \lambda, 0) \cdot V'^T$ with $U' = (u'_{ij})$, $V' = (v'_{ij})$ we obtain

$$\|\widetilde{B} - B\|_f^2 = \mathrm{tr}\left(M^T \cdot M\right) \quad \text{with} \quad M := \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3) - U' \cdot \mathrm{diag}(\lambda, \lambda, 0) \cdot V'^T.$$

By straightforward computation we get

$$\begin{aligned} \mathrm{tr}\left(M^T \cdot M\right) &= (\lambda_1^2 + \lambda_2^2 + \lambda_3^2) + 2\lambda^2 - 2\,\mathrm{tr}\left[\mathrm{diag}(\lambda_1, \lambda_2, \lambda_3) \cdot U' \cdot \mathrm{diag}(\lambda, \lambda, 0) \cdot V'^T\right] = \\ &= (\lambda_1^2 + \lambda_2^2 + \lambda_3^2) + 2\lambda^2 - 2\lambda\left[\lambda_1(u'_{11}v'_{11} + u'_{12}v'_{12}) + \lambda_2(u'_{21}v'_{21} + u'_{22}v'_{22})\right]. \end{aligned}$$

This norm is minimal when the last term in brackets is maximal. We pay attention to $\lambda_1, \lambda_2 > 0$. As U and V are orthogonal, we have

$$(u'_{11}v'_{11} + u'_{12}v'_{12}) \leq 1 \quad \text{and} \quad (u'_{21}v'_{21} + u'_{22}v'_{22}) \leq 1,$$

because each sum can be seen as the scalar product of the top views of the i-th row vectors $\mathbf{u}'_i$ of U' and $\mathbf{v}'_i$ of V', $i = 1, 2$. These sums are maximal if these top views are still unit vectors and $\mathbf{u}'_i = \mathbf{v}'_i$. This implies

$$U' = V' = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$A = U' \cdot \mathrm{diag}(\lambda, \lambda, 0) \cdot V'^T = \mathrm{diag}(\lambda, \lambda, 0).$$

Hence

$$\|\widetilde{B} - B\|_f^2 = (\lambda_1^2 + \lambda_2^2 + \lambda_3^2) + 2\lambda^2 - 2\lambda(\lambda_1 + \lambda_2) = (\lambda_1 - \lambda)^2 + (\lambda_2 - \lambda)^2 + \lambda_3^2,$$

and this is minimal for the arithmetic mean $\lambda = \frac{1}{2}(\lambda_1 + \lambda_2)$ of the greatest two values. □

The factorization $B = S{\cdot}R$ according to Theorem 2 reveals already the relative position of the two camera frames. Therefore we need

Theorem 6 *The factorization of the essential matrix* $B = U{\cdot}D{\cdot}V^T$, $D = \mathrm{diag}(\lambda, \lambda, 0)$, *into the skew-symmetric matrix* S *and the orthogonal matrix* R *reads:*

$$S = \pm U \cdot R_+ \cdot D \cdot U^T, \quad R = \pm U \cdot R_+^T \cdot V^T \quad \text{where} \quad R_+ = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{14}$$

Proof: It is sufficient to factorize the product of the first two matrices of $B = U{\cdot}D{\cdot}V^T$ by

$$U \cdot D = S \cdot R',$$

because this implies immediately

$$B = S \cdot (R'{\cdot}V^T), \ \text{ i.e., } \ R = R'{\cdot}V^T.$$

We again focus on the affine 3D transformations which are represented by the involved matrices:

Figure 12: Epipolar lines. The two epipoles are defined by the row- and the column-nullspace of the essential matrix B

Any skew-symmetric matrix S represents the commutative product of an orthogonal projection with a 90°-rotation and a scaling (see Fig. 8). On the other hand, $U\cdot D$ represents the product of the orthogonal projection parallel to the $\overline{z}$-axis, the scaling with factor λ and the isometry U which brings the $\overline{x}\,\overline{y}$-plane into the image space of $U\cdot D$.

Let R_+ be the matrix representing the 90°-rotation about the $\overline{z}$-axis. Then R_+ is of the form given in Theorem 6, and the product $R_+\cdot D = D\cdot R_+$ is skew-symmetric. This skew-symmetry is preserved under transformation with U while the $\overline{x}\,\overline{y}$-image-plane is transformed into the correct position. This gives the solutions

$$S = \pm U\cdot R_+\cdot D\cdot U^T \quad \text{and} \quad R' = \pm U\cdot R_+^T \quad \text{with} \quad S\cdot R' = U\cdot D.$$

For the following reason there are not more than two different factorizations of the required type: As matrix $B = U\cdot D\cdot V^T$ represents a scaled orthogonal axonometry, the column vectors are the images of an orthonormal frame. We know from Descriptive Geometry that apart from translations parallel to the rays of sight there are exactly two different triples of pairwise orthogonal axes with images in direction of the given column vectors. The two triples are mirror images from each other. So, Theorem 6 gives all possible factorizations. □

There are *critical configurations* where the specified reference points do not determine the epipoles uniquely. This is happens, e.g., when only coplanar 3D points are chosen as reference points, because their images $\mathbf{x}'_i \mapsto \mathbf{x}''_i$ determine a collinear transformation $\kappa\colon \pi' \to \pi''$, and any pair of corresponding points $\mathbf{s}', \mathbf{s}''$ would serve as epipoles (compare Fig. 9). Besides, there are also critical cases related to quadrics. For details see, e.g., [14, 15, 5]).

5. The algorithm

We summarize: The numerical reconstruction of two calibrated images with the aid of any computer algebra system (e.g., Maple) consists of the following six steps:

(1) Specify $n > 7$ pairs $(\mathbf{x}'_i, \mathbf{x}''_i)$, $i = 1, \ldots, n$, of corresponding points (cf. Fig. 11) under avoidance of critical configurations.

(2) Set up the homogeneous linear system of equations (6) for the unknown essential matrix B. The optimal solution $\widetilde{B}$ is an eigenvector of the smallest eigenvalue of $A^T\cdot A$, when A denotes the coefficient matrix of this system.

(3) Based on the singular value decomposition of $\widetilde{B}$ compute the closest rank 2 matrix B with two equal singular values according to Theorem 5. This defines the projectivity between epipolar lines (Fig. 12).

(4) Factorize B into the product $S \cdot R$ of a skew-symmetric matrix S and an orthogonal R according to Theorem 6. This determines the relative position of the two camera frames — up to the scale.

(5) In one of the camera frames compute the approximate point $\mathbf{s}_i$ of intersection between corresponding rays $\mathbf{c}_1 \vee \mathbf{x}'_i$ and $\mathbf{c}_2 \vee \mathbf{x}''_i$, $i = 1, 2, \ldots$ (Fig. 13).

(6) Transform the recovered point coordinates into any world coordinate frame.

Figs. 10, 11 and 12 show an example with the computed epipolar lines and epipoles. The recovered object is displayed in Fig. 14.

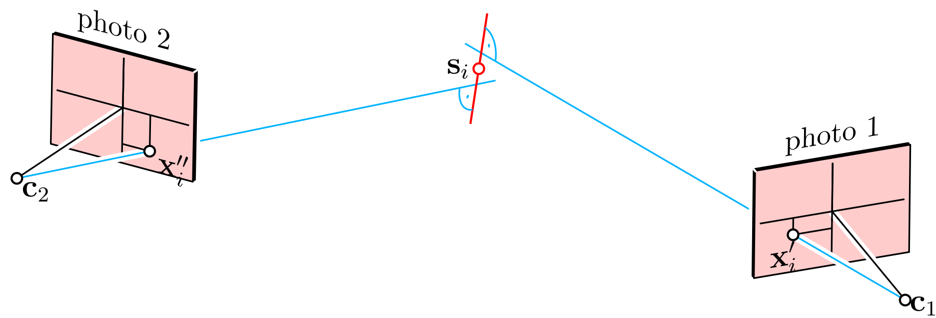

Figure 13: Approximating the point of intersection between corresponding rays

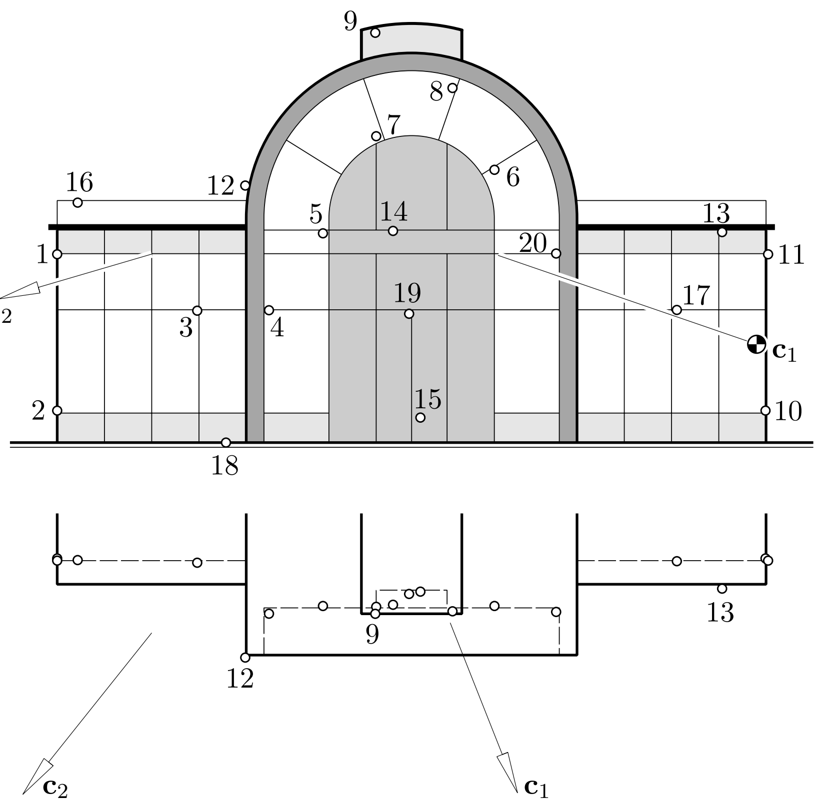

Figure 14: The result of the reconstruction — in top view and front view

References

[1] H. Brauner: *Lineare Abbildungen aus euklidischen Räumen*. Beitr. Algebra Geom. **21**, 5–26 (1986).

[2] A. Dür: *An Algebraic Equation for the Central Projection*. J. Geometry Graphics **7**, 137–143 (2003).

[3] O. Faugeras: *Three-Dimensional Computer Vision. A Geometric Viewpoint*. MIT Press, Cambridge, Mass., 1996.

[4] O. Faugeras, Q.-T. Luong: *The Geometry of Multiple Images*. MIT Press, Cambridge, Mass., 2001.

[5] R. Hartley, A. Zisserman: *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[6] V. Havel: *O rozkladu singulárních lineárních transformací*. Časopis Pést. Mat. **85**, 439–447 (1960).

[7] H. Havlicek: *On the Matrices of Central Linear Mappings*. Math. Bohem. **121**, 151–156 (1996).

[8] M. Hoffmann, P. Yiu: *Moving Central Axonometric Reference Systems*. J. Geometry Graphics **9**, 133–140, (2005).

[9] E. Kruppa: *Zur achsonometrischen Methode der darstellenden Geometrie*. Sitzungsber., Abt. II, österr. Akad. Wiss., Math.-Naturw. Kl. **119**, 487–506 (1910).

[10] C.L. Lawson, R.J. Hanson: *Solving Least Squares Problems*. Siam, 1995, 1st ed.: Prentice-Hall, Inc., Englewood Cliffs, NJ, 1974.

[11] H. Stachel: *Zur Kennzeichnung der Zentralprojektionen nach H. Havlicek*. Sitzungsber., Abt. II, österr. Akad. Wiss., Math.-Naturw. Kl. **204**, 33–46 (1995).

[12] H. Stachel: *On Arne Dür's Equation Concerning Central Axonometries*. J. Geometry Graphics **8**, 215–224 (2004).

[13] J. Szabó, H. Stachel, H. Vogel: *Ein Satz über die Zentralaxonometrie*. Sitzungsber., Abt. II, österr. Akad. Wiss., Math.-Naturw. Kl. **203**, 3–11 (1994).

[14] J. Tschupik, F. Hohenberg: *Die geometrische Grundlagen der Photogrammetrie*. In Jordan, Eggert, Kneissl (eds.): *Handbuch der Vermessungskunde III a/3,* 10. Aufl., Metzlersche Verlagsbuchhandlung, Stuttart 1972, 2235–2295.

[15] Yi Ma, St. Soatto, J. Košecká, S.Sh. Sastry: *An Invitation to 3-D Vision*. Springer-Verlag, New York 2004.

Received August 7, 2006; final form December 29, 2006