"Dunarea de Jos" University, Galati
Graphics, Mechanisms and Tolerances Department

THE 2nd INTERNATIONAL CONFERENCE ON ENGINEERING GRAPHICS AND DESIGN
- ICEGD, 2007 -

June 7 – 10, 2007, Galati, ROMANIA

THE GEOMETRY BEHIND THE NUMERICAL RECONSTRUCTION OF TWO PHOTOS

Hellmuth STACHEL

Abstract: *The geometry of pairs of linear images has been a standard topic of Descriptive Geometry for more than 100 years [2]. During the last twenty years great progress has been made within the field of Computer Vision [3]. Previously graphical or mechanical methods of reconstruction have been replaced by numerical methods. This paper will explain to geometers how to recover metrical data from two images using a computer algebra system. Not the presented results are new, but the way how they are deduced by geometric reasoning.*
Key words: *Descriptive Geometry, multiple images, two-views system, essential matrix, Computer Vision*

1. INTRODUCTION

The basic term in this paper is the *central projection* with *center* $\mathbf{z}$ and *image plane* π, which maps any point $\mathbf{x} \neq \mathbf{z}$ onto the point of intersection $\mathbf{x}^c = (\mathbf{z} \vee \mathbf{x}) \cap \pi$. This is the geometric idealization of the photographic mapping with $\mathbf{z}$ as the focal center of the lenses and π as the plane carrying the CCD sensor. The pedal point of $\mathbf{z}$ with respect to π is the *principal point* $\mathbf{h}$; the distance $d = \|\mathbf{z} - \mathbf{h}\|$ is the *focal length*. The obtained image is called a *central* or *linear perspective*.

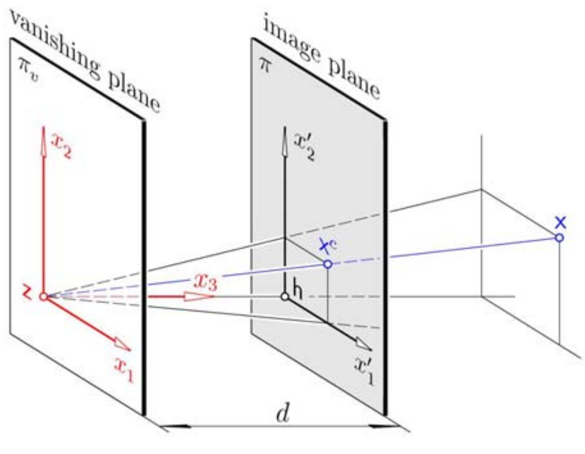

Fig. 1. Central projection and the camera frame

Each central projection or photographic mapping is connected with a particular coordinate frame in space, the *camera frame*. Its origin is placed at the center $\mathbf{z}$, the principal ray $\mathbf{z} \vee \mathbf{h}$ is the x_3-axis. The main directions in the photosensitive plane serve as x_1- and x_2-axis. These two coordinate axes span the vanishing plane π_v of this central projection.

For any given central photo the position of the principal point $\mathbf{h}$ and the focal distance *d* (adjusted to the scale of the photo) define the *intrinsic calibration parameters*. A photo with known intrinsic calibration parameters is called *calibrated*. In this case the position of the camera frame relative to the photo is known. This determines – up to a rigid spatial motion – the bundle of rays $\mathbf{z} \vee \mathbf{x}^c$ in the *viewing situation*, i.e., in the moment when the photo was made.

On the other hand, the parameters which define position of the camera frame with respect to the world coordinate frame are called *extrinsic calibration parameters*. In the sequel we assume that two calibrated photos of the same object are given, and our goal is to recover the metrical data of the depicted object. A classical theorem originating from the Bavarian geometer S. Finsterwalder (1899) states:

Theorem 1 (Fundamental Theorem of Photogrammetry): *From two calibrated images the depicted object can be reconstructed up to a similarity.*

2. THE EPIPOLAR CONSTRAINT

Let two central projections be given with centers $\mathbf{z}_1$, $\mathbf{z}_2$ and image planes π_1 and π_2, respectively. We presuppose $\mathbf{z}_1 \neq \mathbf{z}_2$.

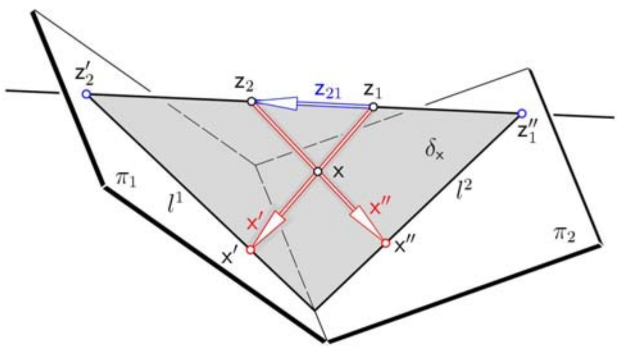

Fig. 2. A pair of perspectives in the viewing situation. $\mathbf{z}_2'$ and $\mathbf{z}_1''$ are the epipoles, $\mathbf{x}'$ and $\mathbf{x}''$ are corresponding images

Any space point $\mathbf{x}$ different from the two centers has two images $\mathbf{x}'$ and $\mathbf{x}''$. We call these two images of $\mathbf{x}$ *corresponding* (Fig. 2), and this remains after the two perspectives are scaled and placed somewhere else thus destroying the viewing situation. This is called a *two-views system*.

The basic geometric property of two-views systems results from the fact that for all space points $\mathbf{x}$ which are not aligned with the two centers, the two rays of sight $\mathbf{z}_1 \vee \mathbf{x}$ and $\mathbf{z}_2 \vee \mathbf{x}$ are coplanar and they span a plane $\delta_\mathbf{x}$ which in both images is depicted in an edge view (see Fig. 2). In the viewing situation the images l^1 and l^2 of these planes $\delta_\mathbf{x}$ constitute two *perspective* line pencils in π_1 and π_2. After the displacement of the two images the pencils remain mutually *projective*. These lines are called *epipolar*. The centers $\mathbf{z}_2{}'$ and $\mathbf{z}_1{}''$ of the two pencils are the *epipoles*. The projectivity between the two pencils is called *epipolar constraint*. We summarize:

Theorem 2: *For any pair of linear perspectives of a scene there is a projectivity between two particular line pencils such that two image points* $\mathbf{x}'$ *and* $\mathbf{x}''$ *are corresponding if and only if they are located on corresponding epipolar lines.*

For the numerical reconstruction we must express the epipolar constraint in an analytical way:

For this purpose we specify the coordinates. For each image point $\mathbf{x}'$ or $\mathbf{x}''$ we use its 3D coordinates with respect to the corresponding camera frame as homogeneous 2D coordinates and we denote this 3-vector again with the same symbol. According to Fig. 2 the three vectors $\mathbf{z}_{21} := \mathbf{z}_2 - \mathbf{z}_1$, $\mathbf{x}'$ and $\mathbf{x}''$ 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

$$\mathbf{x}_1 := \mathbf{z}_{21} + \mathbf{R}.\mathbf{x}_2$$

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

$$0 = \det(\mathbf{x}', \mathbf{z}_{21}, \mathbf{z}_{21} + \mathbf{R}.\mathbf{x}'') = \mathbf{x}'.(\mathbf{z}_{21} \times \mathbf{R}.\mathbf{x}'') .$$

We may replace the cross product with $\mathbf{z}_{21}$ by the product of $\mathbf{x}''$ with a skew-symmetric matrix $\mathbf{S}$ with the coordinates of $\mathbf{z}_{21}$ as entries, i.e.,

$$\mathbf{x}'.(\mathbf{z}_{21} \times \mathbf{R}.\mathbf{x}'') = \mathbf{x}'.(\mathbf{S}.\mathbf{R}.\mathbf{x}'') .$$

Now we write the dot product in matrix form and end up with

Theorem 3: *When for two calibrated photos each image point is represented by its coordinate vector with respect to the camera frame, then the two points* $\mathbf{x}'$ *and* $\mathbf{x}''$ *are corresponding if and only if*

$$\mathbf{x}'^{\mathrm{T}}.\,\mathbf{B}.\mathbf{x}'' = 0 \ \textit{ with the essential matrix } \mathbf{B} = \mathbf{S}.\mathbf{R} .$$

Here $\mathbf{S}$ *is skew-symmetric, i.e.,* $\mathbf{S}^{\mathrm{T}} = -\mathbf{S}$, *and* $\mathbf{R}$ *is orthogonal, i.e.,* $\mathbf{R}^{\mathrm{T}} = \mathbf{R}^{-1}$.

This *essential matrix* $\mathbf{B}$ is the coefficient matrix of a bilinear form and contains nine entries. Its rank is < 3 since any skew-symmetric 3×3 matrix is singular. However, $\mathbf{B}$ is even more special.

Fig. 3. Step 1) Specify pairs of corresponding points and determine their coordinates with respect to the camera frames

Theorem 4: *The essential matrix* $\mathbf{B}$ *of two calibrated photos has two equal singular values.*

Proof: A result of Linear Algebra says that each matrix $\mathbf{B}$ can be expressed by a product $\mathbf{U}^{\mathrm{T}}.\mathbf{D}.\mathbf{V}$ with a diagonal matrix $\mathbf{D}$ of the same size and two orthogonal matrices $\mathbf{U}$ and $\mathbf{V}$. All entries in $\mathbf{D}$ are ≥ 0. The positive digonal elements are called *singular values* or *principal distortions* of $\mathbf{B}$. They are uniquely defined, and their computation is a standard routine for computer-algebra systems. E.g., the related command in Maple reads

```
sing_vals := evalf(Svd(B,U,V)).
```

The singular values of $\mathbf{B} = \mathbf{S}.\mathbf{R}$ are identical with that of the skew-symmetric $\mathbf{S}$, since $\mathbf{R}$ is orthogonal. And they are invariant against the coordinate transformation from $\mathbf{x}$ to $\tilde{\mathbf{x}} = \mathbf{W}.\mathbf{x}$ with a direct orthogonal matrix $\mathbf{W}$, since

$$\mathbf{W}.(\mathbf{S}.\mathbf{x}) = \mathbf{W}.\mathbf{S}.\mathbf{W}^{\mathrm{T}}.\tilde{\mathbf{x}} = \tilde{\mathbf{S}}.\tilde{\mathbf{x}} \text{ , hence } \tilde{\mathbf{S}} = \mathbf{W}.\mathbf{S}.\mathbf{W}^{\mathrm{T}} .$$

On the other hand

$$\tilde{\mathbf{S}}.\tilde{\mathbf{x}} = \mathbf{W}.(\mathbf{z}_{21} \times \mathbf{x}) = (\mathbf{W}.\mathbf{z}_{21}) \times (\mathbf{W}.\mathbf{x}) = (\tilde{\mathbf{z}}_{21} \times \tilde{\mathbf{x}}) .$$

This reveals that the entries in $\tilde{\mathbf{S}}$ are the transformed coordinates of $\mathbf{z}_{21}$. We may assume $\tilde{\mathbf{z}}_{21} = (0,0,z)^{\mathrm{T}}$ with $z = \|\mathbf{z}_{21}\| \neq 0$. Thus we obtain

$$\tilde{\mathbf{S}} = \begin{pmatrix} 0 & -z & 0 \\ z & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} z & 0 & 0 \\ 0 & z & 0 \\ 0 & 0 & 0 \end{pmatrix} . \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{D}.\mathbf{R}_3 .$$

$\mathbf{R}_3$ is orthogonal and represents the rotation about the x_3-axis through 90^0. The matrix $\mathbf{D}$ contains the singular values (z,z) of $\tilde{\mathbf{S}}$, of $\mathbf{S}$ and $\mathbf{B}$.

Conversely, if $\mathbf{B} = \mathbf{U}^{\mathrm{T}}.\mathbf{D}.\mathbf{V}$ is given, then $\mathbf{D}.\mathbf{R}_3$ is skew-symmetric as well as the matrix $\mathbf{U}^{\mathrm{T}}.\mathbf{D}.\mathbf{R}_3.\mathbf{U}$, which is obtained by transformation with $\mathbf{U}$. This leads to the factorization

$$\begin{aligned} \mathbf{B} &= \mathbf{U}^{\mathrm{T}}.\mathbf{D}.(\mathbf{R}_3.\mathbf{U}.\mathbf{U}^{\mathrm{T}}.\mathbf{R}_3^{\mathrm{T}}).\mathbf{V} \\ &= (\mathbf{U}^{\mathrm{T}}.\mathbf{D}.\mathbf{R}_3.\mathbf{U}).(\mathbf{U}^{\mathrm{T}}.\mathbf{R}_3^{\mathrm{T}}.\mathbf{V}) = \mathbf{S}.\mathbf{R} \end{aligned}$$

Theorem 5: *If $\mathbf{B} = \mathbf{U}^{\mathrm{T}}.\mathbf{D}.\mathbf{V}$ is the essential matrix for two calibrated photos, then the decomposition of $\mathbf{B}$ into the product $\mathbf{S}.\mathbf{R}$ of a skew-symmetric and an orthogonal matrix reads $\mathbf{S} = \pm\mathbf{U}^{\mathrm{T}}.\mathbf{D}.\mathbf{R}_3.\mathbf{U}$ and $\mathbf{R} = \pm\mathbf{U}^{\mathrm{T}}.\mathbf{R}_3^{\mathrm{T}}.\mathbf{V}$.*

It is proved in [3] or [1] that this is the only possible factorization of this type. It is important to realize that the two factors $\mathbf{S}$ and $\mathbf{R}$ define the relative position between the two camera frames up to the distance $\|\mathbf{z}_2 - \mathbf{z}_1\|$. Hence we can recover the viewing situation and reconstruct the depicted object by intersecting corresponding rays $\mathbf{z}_1 \vee \mathbf{x}_i'$ and $\mathbf{z}_2 \vee \mathbf{x}_i''$. Any variation of the distance $\|\mathbf{z}_2 - \mathbf{z}_1\|$ acts as a similarity on the recovered object.

For the computation of the essential matrix we use the coordinates of $n > 7$ pairs $(\mathbf{x}_i', \mathbf{x}_i'')$ of corresponding points and apply least square methods to rule out the imprecisions of the measurement. This is carried out in two steps: We first compute the optimal solution $\tilde{\mathbf{B}}$ of the homogeneous system of equations $\mathbf{x}_i'^{\mathrm{T}}.\mathbf{B}.\mathbf{x}_i'' = 0$ originating from the bilinear relation in Theorem 3. This is an eigenvalue problem (see, e.g., [1]). In order to find in the neighborhood of $\tilde{\mathbf{B}}$ a matrix with two equal singular values, we follow

Theorem 6: *Let $\tilde{\mathbf{B}} = \mathbf{U}^{\mathrm{T}}.\tilde{\mathbf{D}}.\mathbf{V}$ be the singular value decomposition of $\tilde{\mathbf{B}}$ with $\tilde{\mathbf{D}} = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$ obeying $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Then the replacement of $\tilde{\mathbf{D}}$ by $\mathbf{D} = \mathrm{diag}(\lambda, \lambda, 0)$ with $\lambda = (\lambda_1 + \lambda_2)/2$ gives the closest essential matrix $\mathbf{B} = \mathbf{U}^{\mathrm{T}}.\mathbf{D}.\mathbf{V}$, i.e., with minimal Frobenius norm $\|\mathbf{B} - \tilde{\mathbf{B}}\|$.*

For the proof see, e.g., [3] or [1].

3. THE ALGORITHM

Now we can formulate the numerical algorithm for the numerical reconstruction of two calibrated images with the aid of any computer algebra system (e.g., Maple). It consists of the following six steps:

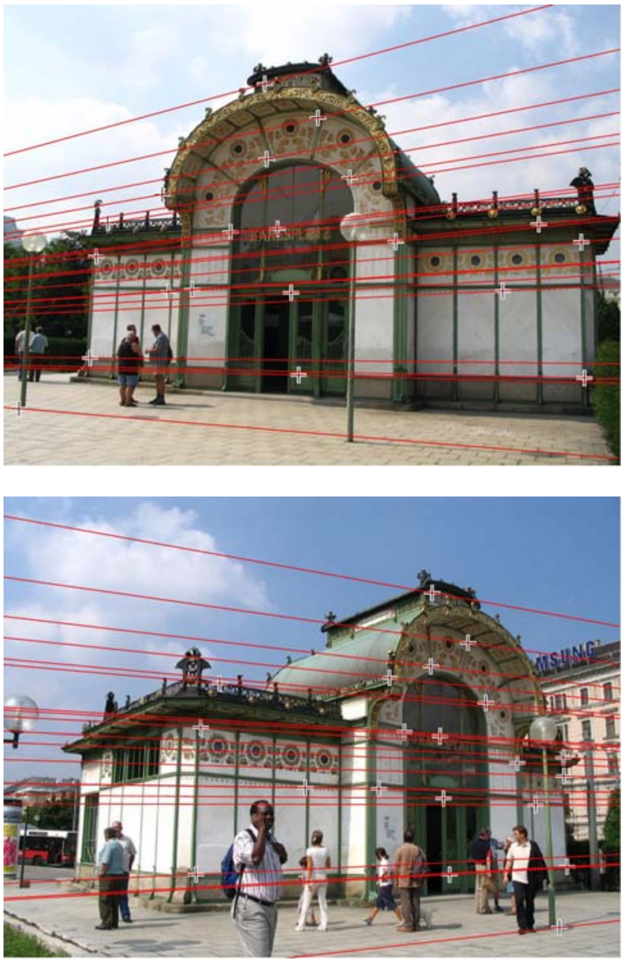

Fig. 4. Step 4) The essential matrix $\mathbf{B} = \mathbf{S}.\mathbf{R}$ determines the relative position of the two camera frames and hence the epipolar lines passing through the epipoles

1) Specify $n > 7$ pairs $(\mathbf{x}_i', \mathbf{x}_i'')$, $i = 1,...,n$, of corresponding points (Fig. 3) and determine their coordinates with respect to the corresponding camera frame by measurements in the photos.
2) Set up the homogeneous linear system of equations $\mathbf{x}_i'^{\mathrm{T}}.\mathbf{B}.\mathbf{x}_i'' = 0$ for the unknown essential matrix $\tilde{\mathbf{B}}$.
3) The optimal solution $\tilde{\mathbf{B}}$ is an eigenvector of the smallest eigenvalue of $\mathbf{A}^{\mathrm{T}}.\mathbf{A}$, when $\mathbf{A}$ denotes the coefficient matrix of the linear system of equations.
4) Based on the singular value decomposition of $\tilde{\mathbf{B}}$, i.e., $\tilde{\mathbf{B}} = \mathbf{U}^{\mathrm{T}}.\tilde{\mathbf{D}}.\mathbf{V}$ with $\mathbf{U}^{\mathrm{T}} = \mathbf{U}^{-1}$, $\tilde{\mathbf{D}} = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$, $\lambda_1 \geq \lambda_2 \geq \lambda_3$, $\mathbf{V}^{\mathrm{T}} = \mathbf{V}^{-1}$, compute the closest rank 2 matrix with two equal singular values. Due to Theorem 6 this is $\mathbf{B} = \mathbf{U}^{\mathrm{T}}.\mathbf{D}.\mathbf{V}$ with the arithmetic mean $\lambda = (\lambda_1 + \lambda_2)/2$ as entry in the diagonal matrix $\mathbf{D} = \mathrm{diag}(\lambda, \lambda, 0)$. Matrix $\mathbf{B}$ defines the projectivity

between epipolar lines (Figs. 2 and 4). The epipole $\mathbf{z}_1"$ solves the equation $\mathbf{B}.\mathbf{x} = \mathbf{0}$, the epipole $\mathbf{z}_2'$ is a solution of the homogeneous system $\mathbf{B}^{\mathrm{T}}.\mathbf{x} = \mathbf{0}$.

5) Factorize $\mathbf{B}$ into the product $\mathbf{S}.\mathbf{R}$ of a skew-symmetric matrix $\mathbf{B}$ and an orthogonal $\mathbf{R}$ according to Theorem 5. This determines the relative position of the two camera frames – up to the scale.
6) In one of the camera frames compute the approximate point $\mathbf{s}_i$ of intersection between corresponding rays $\mathbf{z}_1 \vee \mathbf{x}_i'$ and $\mathbf{z}_2 \vee \mathbf{x}_i"$ for $i = 1, \ldots, n$ via the common intersecting normal line (Fig. 5) and inspect the square sum of the shortest distances as a measure for the obtained precision.
7) Transform the recovered point coordinates of the scene into any world coordinate frame.

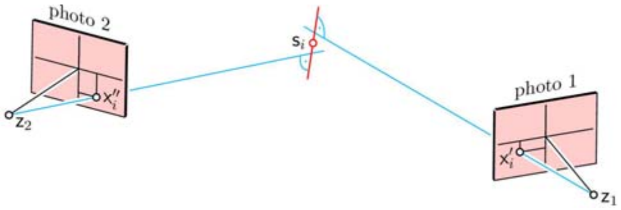

Fig. 5. Step 6) Compute the approximate point $\mathbf{s}_i$ of intersection between corresponding rays $\mathbf{z}_1 \vee \mathbf{x}_i'$ and $\mathbf{z}_2 \vee \mathbf{x}_i"$

Figs. 3-4 show an example with the computed epipolar lines and epipoles. The recovered object is displayed in Fig. 6.

ICEGD 2007
Code and name of the conference section
University *"Dunarea de Jos"* of Galati, Faculty of Mechanical Engineering - **Graphics, Mechanisms and Tolerances Department,** Domneasca Str., 111, Postcode: 800201

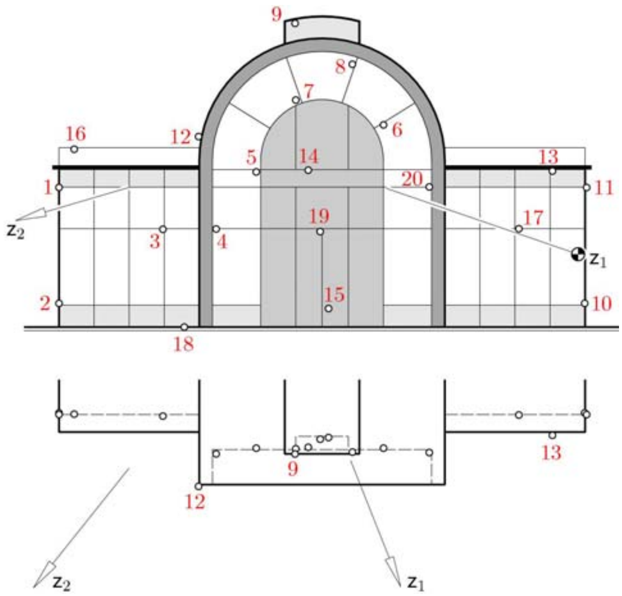

Fig. 6. The reconstructed object, the historical `Stadtbahn' station Karlsplatz in Vienna (Otto Wagner, 1897)

4. REFERENCES

[1] Stachel, H. (2006). *Descriptive Geometry Meets Computer Vision - The Geometry of Two Images.* J. Geometry Graphics 10 (2), 137-153, ISSN 1 433-8157.
[2] Tschupik, J., Hohenberg, F. (1972). Die geometrischen Grundlagen der Photogrammetrie. In: Handbuch der Vermessungskunde III a/3, Jordan, Eggert, Kneissl (Eds.), 2235–2295, 10th ed., Metzlersche Verlagsbuchhandlung, Stuttart.
[3] Yi Ma, Soatto, St., Kosecká, J., Sastry, S.Sh. (2004). *An Invitation to 3-D Vision,* Springer-Verlag, ISBN 0-387-00893-4, New York

Author: Hellmuth STACHEL, professor, Vienna University of Technology, Institute of Discrete Mathematics and Geometry, E-mail: stachel@dmg.tuwien.ac.at, Phone: +43 1 58801 11320.