SCENE CONSTRAINTS FOR DIRECT SINGLE IMAGE ORIENTATION WITH SELFDIAGNOSIS

Mirko Appel[a] and Wolfgang Förstner[b]

[a] Siemens AG, Corporate Technology, Otto-Hahn-Ring 6, 81730 München, Germany – mirko.appel@siemens.com*
[b] Institut für Photogrammetrie, Universität Bonn, Nussallee 15, 53115 Bonn, Germany – wf@ipb.uni-bonn.de

KEY WORDS: Orientation from Points and Lines, Industrial Application, Projective Geometry, Maximum Likelihood Estimation, 3D Image Map Registration

ABSTRACT

In this paper we present a new method for single image orientation using an orthographic drawing or map of the scene. Environments which are dominated by man made objects, such as industrial facilities or urban scenes, are very rich of vertical and horizontal structures. These scene constraints reflect in symbols in an associated drawing. For example, vertical lines in the scene are usually marked as points in a drawing. The resulting orientation may be used in augmented reality systems or for initiating a subsequent bundle adjustment of all available images.
In this paper we propose to use such scene constraints taken from a drawing to estimate the camera orientation. We use observed vertical lines, horizontal lines, and points to estimate the projection matrix P of the image. We describe the constraints in terms of projective geometry which makes them straightforward and very transparent. In contrast to the work of (Bondyfalat et al., 2001), we give a direct solution for P without using the fundamental matrix between image and map as we do not need parallelity constraints between lines in a vertical plane other than for horizontal lines, nor observed perpendicular lines.
We present both a direct solution for P and a statistically optimal, iterative solution, which takes the uncertainties of the contraints and the observations in the image and the drawing into account. It is a simplifying modification of the eigenvalue method of (Matei and Meer, 1997). The method allows to evaluate the results statistically, namely to verify the used projection model and the assumed statistical properties of the measured image and map quantities and to validate the achieved accuracy of the estimated projection matrix P.
To demonstrate the feasibility of the approach, we present results of the application of our method to both synthetic data and real scenes in industrial environment. Statistical tests show the performance and prove the rigour of the new method.

1 INTRODUCTION

Many tasks in computer vision and photogrammetry require prior estimation of a camera's orientation and calibration. This problem has been well investigated in the past by many researchers (Faugeras, 1993, Kanatani, 1996, Klette et al., 1998, Faugeras and Luong, 2001). Most of these methods use point and line correspondences between the images and/or calibration patterns to estimate the camera's intrinsic and extrinsic parameters. However, camera calibration may be more reliable and easier to carry out if further scene constraints are taken into account. Scenes which are dominated by man made objects are usually very rich of such constraints. For example, urban scenes or scenes in industrial environment contain lots of vertical and horizontal structures (see Fig. 1). Many works in computer vision and photogrammetry literature exploit these constraints by using vanishing points for recovery of camera orientation (Caprile and Torre, 1990, Wang and Tsai, 1991, Youcai and Haralick, 1999, van den Heuvel, 1999). Here, we propose to take advantage of horizontal and vertical structures in conjunction with a map or drawing of the scene. It is important to note that very often such maps or drawings are readily available. Drawings are with no doubt the most important and commonly used documents in many industries. These documents are created during the design process, and they are used and completed by builders. Furthermore, drawings are referred to on a daily basis for maintenance and update of buildings and facilities. It is therefore quite advantageous to use these documents. In practice, many have taken advantage of such documents to find some reference points in order to register virtual and real world coordinates.

In this paper we aim at providing methods for direct single view orientation using these commonly available documents. We use vertical lines, horizontal lines, and points to estimate the projection matrix P. We describe the constraints in terms of projective geometry which makes them straightforward and very transparent. In contrast to the work of (Bondyfalat et al., 2001), we give a direct solution for P without using the fundamental matrix between image and map as we do not need parallelity constraints between lines in a vertical plane other than for horizontal lines.

We present both a direct solution for P and a statistically optimal, iterative solution, which takes the uncertainties of the contraints and the observations in the image into account. It is a simplifying modification of the eigenvalue method of (Matei and Meer, 1997). The method allows to evaluate the results statistically, namely to verify the used projection model and the assumed statistical properties of

*This work was done while M. Appel was with the Institut für Photogrammetrie, Universität Bonn

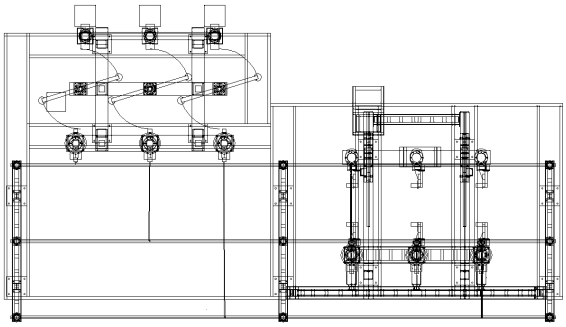

Figure 1: Image of a high voltage switch gear and the associated drawing. Many horizontal and vertical structures occur in the image, which correspond to lines and points in the drawing.

the measured image and map quantities and to validate the achieved accuracy of the estimated projection matrix P.

The result of such an image-drawing orientation may be used in augmented reality systems or just as approximate values for a subsequent rigorous bundle adjustment using all available images.

The remainder of this paper is organized as follows: After introducing notation, geometric relations and statistical essentials in section 2, we describe the scene constraints in terms of projective geometry in section 3. In section 4 we provide both an algebraic solution for P and a statistically optimal one, along with statistical tools for self-diagnosis. Experiments on synthetic and real data are presented in section 5. We finally conclude in section 6.

2 MATHEMATICAL FOUNDATIONS

First, we briefly describe the represention of geometric entities, the assumed representation for uncertain geometric entities, basic error propagation, estimation and testing techniques used later in the paper.

Notation: We denote coordinate vectors of planar geometric objects with small bold face letters, e. g. $\boldsymbol{x}$, in 3D space with capital bold face letters, e. g. $\boldsymbol{X}$. Vectors and matrices are denoted with slanted letters, thus $\boldsymbol{x}$ or $\boldsymbol{R}$. Homogeneous vectors and matrices are denoted with upright letters, e. g. $\mathbf{x}$ or $\mathbf{A}$. We use the skew matrix $\mathsf{S}(\mathbf{x}) = [\mathbf{x}]_\times$ of a 3-vector $\mathbf{x}$ in order to represent the cross product by $\mathbf{a} \times \mathbf{b} = \mathsf{S}(\mathbf{a})\mathbf{b} = -\mathbf{b} \times \mathbf{a} = -\mathsf{S}(\mathbf{b})\mathbf{a}$. We will use the vec-operator collecting the columns of a matrix into a vector, thus $\mathrm{vec}\mathsf{A} = \mathrm{vec}(\boldsymbol{a}_1, \boldsymbol{a}_2, ...) = (\boldsymbol{a}_1^\mathsf{T}, \boldsymbol{a}_2^\mathsf{T}, ...)^\mathsf{T}$, and use the relation

$$\mathrm{vec}(\mathsf{ABC}) = (\mathsf{A} \otimes \mathsf{C}^\mathsf{T})\mathrm{vec}(\mathsf{B}^\mathsf{T}) = (\mathsf{C}^\mathsf{T} \otimes \mathsf{A})\mathrm{vec}\mathsf{B} \quad (1)$$

with the Kronecker product $\mathsf{A} \otimes \mathsf{B} = \{a_{ij}\mathsf{B}\}$.

2.1 Geometry

We mainly follow the representation of projective geometry as introduced in (Hartley and Zisserman, 2000). Hence we represent points and lines both in 2D and 3D by homogeneous vectors. We write 2D points as $\mathbf{x} = (u, v, w)^\mathsf{T}$ and 2D lines as $\mathbf{l} = (a, b, c)^\mathsf{T}$. Lines in 2D can easily be constructed from two points by $\mathbf{l} = \mathbf{x} \times \mathbf{y}$ and the incidence constraint of a point and a line is given by $\mathbf{x}^\mathsf{T}\mathbf{l} = 0$. We only need points in 3D, they are represented analogously by $\mathbf{X} = (U, V, W, T)^\mathsf{T}$.

We can write the projection of a 3D point $\mathbf{X}$ into an image point $\mathbf{x}$ image plane as

$$\mathbf{x} = \mathsf{P}\mathbf{X} = \begin{pmatrix} \mathsf{P}_1^\mathsf{T} \\ \mathsf{P}_2^\mathsf{T} \\ \mathsf{P}_3^\mathsf{T} \end{pmatrix} \mathbf{X} = (I_3 \otimes \mathbf{X}^\mathsf{T})\mathbf{p} \quad (2)$$

$$\text{with} \qquad \mathbf{p} = \mathrm{vec}(\mathsf{P}^\mathsf{T}) = (\mathsf{P}_1, \mathsf{P}_2, \mathsf{P}_3)^\mathsf{T},$$

where P is the 3×4 projection matrix, P_i^T its rows, and $\mathbf{p} = \mathrm{vec}(\mathsf{P}^\mathsf{T})$ its vector version. Note that due to its homogeneity, only 11 of its elements are independent.

Analogically, we can write the parallel projection yielding the drawing (cf. fig. 1)

$$\mathbf{x}^d = \mathsf{P}^d\mathbf{X} \qquad \text{with} \qquad \mathsf{P}^d = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

inducing $\boldsymbol{x}^d = (x, y)^\mathsf{T}$.The task is to derive P from observed points $\mathbf{x}$ and $\mathbf{x}^d$ in the image and the drawing and further constraints from the interpretaion of the drawing. Straight lines $\mathbf{l}$ in the image always are assumed to be derived by two measured points $\mathbf{x}$ and $\mathbf{y}$, the same holds for lines in the drawing.

2.2 Statistics

Estimation with linear constraints: All constraints will have the form

$$\mathsf{A}(\mathbf{y})\mathbf{p} = \mathbf{e} \stackrel{!}{=} \mathbf{0}$$

where A is a matrix depending on measurements collected in the vector of observations $\mathbf{y}$, and $\mathbf{p}$ is the vector of the unknown elements of the projection matrix, appearing linear in the constraints.

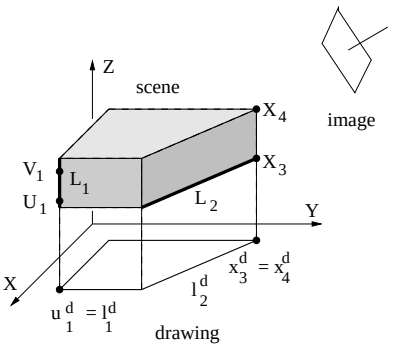

Figure 2: Relation between scene, image and drawing. Vertical line $\mathbf{L}_1$, passing through point $\mathbf{U}_1$, maps to $\mathbf{l}_1^d = \mathbf{u}_1^d$. Horizontal line $\mathbf{L}_2$ has the same vanishing point as its projection $\mathbf{l}_2^d$.

Minimizing the Euclidean distance of $\mathbf{e}$ and $\mathbf{0}$

$$\mathbf{e}^\mathsf{T}\mathbf{e} = \mathbf{p}^\mathsf{T}\mathsf{A}^\mathsf{T}\mathsf{A}\mathbf{p} \qquad \text{under} \qquad |\mathbf{p}| = 1,$$

leads to the well known algebraic solution

$$\mathsf{A}^\mathsf{T}\mathsf{A}\widehat{\mathbf{p}} = \lambda\widehat{\mathbf{p}},$$

namely the right eigenvalue of A.

The statistically optimal solution instead of the Euclidean distance minimizes the Mahalanobis distance of $\mathbf{e}$ and $\mathbf{0}$

$$\mathbf{e}^\mathsf{T}\mathbf{\Sigma}_{ee}^{-1}\mathbf{e} = \mathbf{p}^\mathsf{T}\mathsf{A}^\mathsf{T}\mathbf{\Sigma}_{ee}^{-1}\mathsf{A}\mathbf{p} \qquad \text{under} \qquad |\mathbf{p}| = 1.$$

Here the covariance matrix of the residuals $\underline{\mathbf{e}} = \underline{\mathbf{e}}(\underline{\mathbf{y}}, \mathbf{p})$ can be obtained from error propagation

$$\mathbf{\Sigma}_{ee} = \mathsf{B}(\mathbf{y},\mathbf{p})\mathbf{\Sigma}_{yy}\mathsf{B}^\mathsf{T}(\mathbf{y},\mathbf{p}) \quad \text{with} \quad \frac{\partial \mathbf{e}(\mathbf{y},\mathbf{p})}{\partial \mathbf{y}} = \mathsf{B}(\mathbf{y},\mathbf{p}). \tag{3}$$

It can be shown (see (Förstner, 2001)) that the solution can be obtained iteratively from

$$\left[\mathsf{A}(\widehat{\boldsymbol{y}}^{(\nu)})^\mathsf{T}(\widehat{\mathbf{\Sigma}}_{ee}^{(\nu)})^{-1}\mathsf{A}(\boldsymbol{y})\right]\ \mathbf{p}^{(\nu+1)} = \lambda\mathbf{p}^{(\nu+1)}$$

using

$$\widehat{\mathbf{\Sigma}}_{ee}^{(\nu)} = \left(\mathsf{B}(\widehat{\boldsymbol{y}}^{(\nu)}, \widehat{\mathbf{p}}^{(\nu)})\ \ \mathbf{\Sigma}_{yy}\ \ \mathsf{B}^\mathsf{T}(\widehat{\boldsymbol{y}}^{(\nu)}, \widehat{\mathbf{p}}^{(\nu)})\right),$$

where we need the fitted values $\widehat{\boldsymbol{y}}$ of the observations:

$$\widehat{\boldsymbol{y}}_i^{(\nu)} = \left(\mathsf{I} - \mathbf{\Sigma}_{y_iy_i}\mathsf{B}_i^\mathsf{T}\left(\mathsf{B}_i\mathbf{\Sigma}_{yy}\mathsf{B}_i^\mathsf{T}\right)^{-1}\mathsf{B}_i\right)\mathbf{y}_i,$$

and where the Jacobian $\mathsf{B} = \mathsf{B}(\widehat{\boldsymbol{y}}^{(\nu-1)}, \widehat{\mathbf{p}}^{(\nu-1)})$ has to be evaluated at the fitted values of the previous iteration $(\nu - 1)$.

Selfdiagnosis and performance characteristics: The result can be evaluated based on the optimal value of the Mahalanobis distance

$$\Omega = \widehat{\mathbf{e}}^\mathsf{T}\mathbf{\Sigma}_{ee}^{-1}\widehat{\mathbf{e}} = \widehat{\mathbf{p}}^\mathsf{T}\mathsf{A}^\mathsf{T}\mathbf{\Sigma}_{ee}^{-1}\mathsf{A}\widehat{\mathbf{p}} \sim \chi_R^2,$$

where the Jacobian $\mathsf{A} = \mathsf{A}(\widehat{\mathbf{y}})$ and $\mathbf{\Sigma}_{ee}^{-1}$needs to be evaluated at the fitted values.

In case the assumed model holds, $\underline{\Omega}$ is χ_R^2 distributed with R degrees of freedom, where R is the redundancy of the system:

$$R = G - (U - 1)$$

which is the number of independent constraints minus the number of unknown parameters, here 11.

Assuming the test has not been rejected, we may use the estimated covariance matrix of the estimated parameters

$$\widehat{\mathbf{\Sigma}}_{\widehat{p}\widehat{p}} = \widehat{\sigma}_0^2(\mathsf{A}^\mathsf{T}\mathbf{\Sigma}_{ee}\mathsf{A})^+, \quad \text{with} \quad \widehat{\sigma}_0^2 = \frac{\Omega}{R}, \tag{4}$$

to evaluate the obtained accuracy of the parameters. The calculation of the pseudo inverse can make use of the known nullspace $\widehat{\mathbf{p}}$ of the matrix. In case the geometric constraints hold, one can conclude that the observations have a standard deviation which is larger by a factor of $\widehat{\sigma}_0^2$.

Further, we could use the covariance matrix $\widehat{\mathbf{\Sigma}}_{\widehat{p}\widehat{p}}$ to predict the reprojection error:

$$\widehat{\mathbf{\Sigma}}_{\widehat{x}\widehat{x}} = (\mathsf{I}_3 \otimes \mathbf{X}^\mathsf{T})\widehat{\mathbf{\Sigma}}_{\widehat{p}\widehat{p}}(\mathsf{I}_3 \otimes \mathbf{X}^\mathsf{T})^\mathsf{T} + \widehat{\mathsf{P}}\mathbf{\Sigma}_{XX}\widehat{\mathsf{P}}^\mathsf{T} \tag{5}$$

using (2), $\widehat{\mathbf{x}} = \widehat{\mathsf{P}}\mathbf{X}$, and taking the uncertainty of both $\widehat{\mathsf{P}}$ and $\mathbf{X}$ into account.

Or one could determine the covariance matrix $\widehat{\mathbf{\Sigma}}_{\widehat{Z}\widehat{Z}}$ of the estimated projection center $\widehat{\boldsymbol{Z}} = -\widehat{\mathsf{H}}^{-1}\widehat{\mathbf{h}}$ from $\widehat{\mathsf{P}} = (\widehat{\mathsf{H}}|\widehat{\mathbf{h}}) = (\widehat{\mathsf{H}}| - \widehat{\mathsf{H}}\widehat{\mathbf{Z}})$ (for the derivation cf. appendix):

$$\widehat{\mathbf{\Sigma}}_{\widehat{Z}\widehat{Z}} = \widehat{\mathsf{H}}^{-1}\ (\widehat{\mathbf{Z}}^\mathsf{T} \otimes \mathsf{I}_3)\ \widehat{\mathbf{\Sigma}}_{\widehat{p}\widehat{p}}\ (\widehat{\mathbf{Z}} \otimes \mathsf{I}_3)\ \widehat{\mathsf{H}}^{-\mathsf{T}}.$$

This uncertainty may be compared with some specification coming from the application.

3 SCENE CONSTRAINTS

In the following we derive constraints between measurable quantities in the image on one hand and the map and the unknown projection matrix on the other hand.

3.1 Vertical Lines

We first turn our attention to vertical structures in the scene. Man made objects are very rich of such features, one may think of buildings, all kinds of technical installations, etc.

Observations $\mathbf{l}_i, i = 1, ..., I$ of vertical lines $\mathbf{L}_i$ in the image reflect as a point coordinate in the drawing, say $\boldsymbol{x}_i^d = (x^d, y^d)_i$. Two 3D-points $\mathbf{U}_i$ and $\mathbf{V}_i$ on the same vertical line, therefore, only differ by their third component (cf. fig. 2). Thus, the vertical line is fully characterized by two points

$$\mathbf{U}_i = (X, Y, Z_1, 1)_i^\mathsf{T} \quad \text{and} \quad \mathbf{V}_i = (X, Y, Z_2, 1),^\mathsf{T} \tag{6}$$

where $X_i = x_i^d$ and $Y_i = y_i^d$. The two Z-values Z_1 and Z_2 may be arbitrarily chosen, however, for reasons of numerical stability they should be chosen within the range of the heights, e. g. the minimum and maximum height in the scene.

In general, for the projection of a 3D point on the image, say $\mathbf{x}_i = \mathsf{P}\mathbf{X}_i$, which lies on a line $\mathbf{l}_i$ in the image, $\mathbf{l}_i^\mathsf{T}\mathbf{x}_i = \mathbf{l}_i{'}^\mathsf{T}\mathsf{P}\mathbf{X}_i = 0$ should hold. Hence, two points on a vertical line provide us with two constraints on the projection matrix. We have

$$\begin{pmatrix} \mathbf{l}_i^\mathsf{T}\mathsf{P}\mathbf{U}_i \\ \mathbf{l}_i^\mathsf{T}\mathsf{P}\mathbf{V}_i \end{pmatrix} = \boldsymbol{e}_i \stackrel{!}{=} \mathbf{0} \quad \text{or} \quad \begin{pmatrix} \mathbf{l}_i^\mathsf{T} \otimes \mathbf{U}_i^\mathsf{T} \\ \mathbf{l}_i^\mathsf{T} \otimes \mathbf{V}_i^\mathsf{T} \end{pmatrix} \mathbf{p} = \boldsymbol{e}_i \stackrel{!}{=} \mathbf{0}. \tag{7}$$

Thus, the two residuals $\mathbf{e}_i$ of the constraints should vanish.

Introducing a third point constraint on the same vertical line would not add any information since the line is already fully defined by two points. When adding a third constraint, the rank of the matrix on the left hand side of (7b) would not exceed two.

3.2 Horizontal Lines

Besides vertical lines, horizontal lines are frequently observable features. In this case, observed lines in the image correspond to lines in the drawing.

Let the line in the drawing be defined by two measured points $\mathbf{x}_j^d$ and $\mathbf{y}_j^d$. The point at infinity of that line is $(\boldsymbol{y}_j^{d\mathsf{T}} - \boldsymbol{x}_j^{d\mathsf{T}}, 0)$. As the point at infinity of a line does not depend on its position we obtain the point at infinity of the 3D-line as

$$\mathbf{X}_{\infty j} = \begin{pmatrix} \boldsymbol{y}_j^d - \boldsymbol{x}_j^d \\ \underset{2\times1}{\mathbf{0}} \end{pmatrix} = \begin{pmatrix} R_j \\ S_j \\ 0 \\ 0 \end{pmatrix}.$$

Actually R_j and S_j are the coordinate differences of the two points defining the direction of the line in the drawing.

The observed line $\mathbf{l}_j$ should pass through the image of $\mathbf{X}_{\infty j}$, namely $\mathsf{P}\mathbf{X}_{\infty j}$, known as the vanishing point. Hence, we obtain

$$\mathbf{l}_j^\mathsf{T}\mathsf{P}\mathbf{X}_{\infty j} = \mathbf{e}_j \stackrel{!}{=} 0 \quad \text{or} \quad (\mathbf{l}_j \otimes \mathbf{X}_{\infty j})^\mathsf{T}\mathbf{p} = \mathbf{e}_j \stackrel{!}{=} 0, \tag{8}$$

only constraining the first two columns of P. They do not have any influence on the third and fourth column of P. In fact, by providing a line in the drawing, we only define the direction of this line. We are neither constraining the line's height nor can we use the horizontal position of the line. Therefore, we also may use contours of *horizontal cylinders* with a given direction.

3.3 Observed Points

The third type of scene constraint we consider in this paper is the observation of points in the image which are marked as points in the drawing, which is the classical setup of estimating the projection matrix from points. Let the observation of a point feature in the image $\mathbf{x}_k = (u_k, v_k, w_k)^\mathsf{T}$ be corresponding to a point element in the drawing $\mathbf{X}_k = (U_k, V_k, W_k, T_k)^\mathsf{T}$. Both points are linked by projection such that $\mathbf{x}_k = \mathsf{P}\mathbf{X}_k$ holds. Thus, we obtain the constraints

$$\begin{aligned} \mathbf{x}_k \times \mathsf{P}\mathbf{X}_k &= \mathbf{0} \\ \text{or} \quad \mathsf{S}_{x_k}\mathsf{P}\mathbf{X}_k = -\mathsf{S}_{PX_k}\mathbf{x}_k = \mathbf{f}_k &\stackrel{!}{=} \mathbf{0} \\ \text{or} \quad (\mathsf{S}_{x_k} \otimes \mathbf{X}_k^\mathsf{T})\mathbf{p} = \mathbf{f}_k &\stackrel{!}{=} \mathbf{0}. \end{aligned}$$

Only two of these constraints are independent. As the measured points are finite, the first two constraints are guaranteed to be independent. Thus, we use the two constraints

$$\begin{aligned} \overline{\mathsf{S}}_{x_k}\mathsf{P}\mathbf{X}_k = -\overline{\mathsf{S}}_{PX_k}\mathbf{x}_k = \mathbf{e}_k &\stackrel{!}{=} \mathbf{0} \\ \text{or} \quad (\overline{\mathsf{S}}_{x_k} \otimes \mathbf{X}_k^\mathsf{T})\mathbf{p} = \mathbf{e}_k &\stackrel{!}{=} \mathbf{0} \end{aligned} \tag{9}$$

with

$$\overline{\mathsf{S}}_{x_k} \doteq \begin{pmatrix} \boldsymbol{u}_1^\mathsf{T} \\ \boldsymbol{u}_2^\mathsf{T} \end{pmatrix} \mathsf{S}_{x_k} = \begin{pmatrix} 0 & -w_k & v_k \\ w_k & 0 & -u_k \end{pmatrix}, \tag{10}$$

where $\boldsymbol{u}_i$ is the i-th unit vector.

3.4 Other Constraints

In (Bondyfalat et al., 2001) additional constraints are expoited and used to constrain the fundamental matrix F and the projection matrix: Observing parallel lines in a given plane, leads to constraints which are linear in the elements of F but quadratic in the elements of P. In our scheme we, therefore, cannot include them in the same manner. However, parallel lines which are in the horizontal plane are likely to be contained in the drawing and parallel lines not being horizontal or vertical are not very likely to be present at the object. The same argument holds for observing two orthogonal lines in a given plane. Not using these two types of constraints, therefore, is practically acceptable and gives the way for a direct and an optimal solution of P.

4 ESTIMATION OF P

We propose to use a two step procedure to estimate P similar to (Matei and Meer, 1997). In the first step we solve directly for P using the classical algebraic solution. In the second step we take the uncertainties of all measurements into account in order to obtain a statistically optimal solution.

4.1 Direct Solution

In order to obtain a first estimate for P we integrate all constraints into a system of equations and solve in a least squares sense.

Writing the constraints for $i = 1, \ldots I$ observations of vertical lines, $j = 1, \ldots J$ horizontal lines, and $k = 1, \ldots K$

points in a system of equations, we obtain:

$$\left(\begin{array}{ccc} a_i\mathbf{U}_i^\mathsf{T} & b_i\mathbf{U}_i^\mathsf{T} & c_i\mathbf{U}_i^\mathsf{T} \\ a_i\mathbf{V}_i^\mathsf{T} & b_i\mathbf{V}_i^\mathsf{T} & c_i\mathbf{V}_i^\mathsf{T} \\ \hline a_j\mathbf{X}_{\infty j}^\mathsf{T} & b_j\mathbf{X}_{\infty j}^\mathsf{T} & c_j\mathbf{X}_{\infty j}^\mathsf{T} \\ \hline 0 & -w_k\mathbf{X}_k^\mathsf{T} & v_k\mathbf{X}_k^\mathsf{T} \\ w_k\mathbf{X}_k^\mathsf{T} & 0 & -u_k\mathbf{X}_k^\mathsf{T} \end{array}\right)\left(\begin{array}{c}\mathsf{P}_1\\ \mathsf{P}_2\\ \mathsf{P}_3\end{array}\right) = \left(\begin{array}{c} \mathbf{l}_i^\mathsf{T}\otimes\mathbf{U}_i^\mathsf{T} \\ \mathbf{l}_i^\mathsf{T}\otimes\mathbf{V}_i^\mathsf{T} \\ \hline \mathbf{l}_j^\mathsf{T}\otimes\mathbf{X}_{\infty j}^\mathsf{T} \\ \hline \overline{\mathsf{S}}_{x_k}\otimes\mathbf{X}_k \end{array}\right)\mathbf{p} = \left(\begin{array}{c}\mathbf{e}_i\\ \mathbf{e}_j\\ \mathbf{e}_k\end{array}\right) \stackrel{!}{=} \mathbf{0} \tag{11}$$

the matrix being of size $(2I + J + 2K) \times 12$. The solution which minimizes $\sum_i |\boldsymbol{e}_i|^2 + \sum_j |\boldsymbol{e}_j|^2 + \sum_k |\boldsymbol{e}_k|^2$ is the right eigenvector.

Note that almost any combination of scene constraints may be used in a way that the overall number of constraint is equal to or greater than 11. However, at least two point observations in different heights have to be provided in order to fix the vertical origin and the vertical scaling.

The solution may be changed by arbitrarily weighting the individual constraints, therefore, it definitely is not optimal.

4.2 Uncertainty of Measurements

For achieving an optimal solution we may want to take the uncertainty of the measured entities into account. These are the points $\boldsymbol{x}_i^d$, $\boldsymbol{x}_j^d$, $\boldsymbol{y}_j^d$ and $\mathbf{X}_k$ from the drawing and the points $\boldsymbol{x}_k$ from the image. We, therefore, have to make assumptions about their uncertainty, or – better – derive it empirically.

In our investigation we adopt the following assumptions:

Points in the drawing: We assume the points in the drawing to be measured independently with equal accuracy in both coordinates: Therefore, we assume their homogeneous coordiante vector to be normally distributed with

$$\underline{\mathbf{x}}^{\boldsymbol{d}} \sim N(\boldsymbol{\mu}_{x^d}, \boldsymbol{\Sigma}_{x^dx^d}), \text{where} \boldsymbol{\Sigma}_{x^dx^d} = \sigma_{x^d}^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

In case we know the height of a point from the drawing, we assume it to be defined with the same accuracy. Therefore, the uncertainty of the 3D-points taken from the drawing are characterized by:

$$\underline{\mathbf{X}}^d \sim N(\boldsymbol{\mu}_{X^d}, \boldsymbol{\Sigma}_{X^dX^d}),$$

$$\boldsymbol{\Sigma}_{X^dX^d} = \sigma_{X^d}^2 \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&0 \end{pmatrix} \text{ or } = \sigma_{X^d}^2 \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \end{pmatrix}$$

with $\sigma_{x^d} = \sigma_{X^d}$. The second covariance matrix is taken for points with fixed heights.

The points in the image are measured with a different accuracy. We assume the same simple structure of the distribution:

$$\underline{\mathbf{x}} \sim N(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}), \text{ where } \boldsymbol{\Sigma}_{xx} = \sigma_x^2 \begin{pmatrix} 1&0&0 \\ 0&1&0 \\ 0&0&0 \end{pmatrix}. \tag{12}$$

2D-lines: Original measurements in the image as well as in the drawing consist of points only, points at infinity and lines are treated as derived entities.

Following standard error propagation methods, we obtain the uncertainty of an image (or a drawing) line $\mathbf{l} = \mathbf{x}\times\mathbf{y} = \mathsf{S}_x\mathbf{y} = -\mathsf{S}_y\mathbf{x}$ by

$$\begin{aligned} \boldsymbol{\Sigma}_{ll} &= \frac{\partial \mathbf{l}}{\partial \mathbf{y}}\boldsymbol{\Sigma}_{yy}\left(\frac{\partial \mathbf{l}}{\partial \mathbf{y}}\right)^\mathsf{T} + \frac{\partial \mathbf{l}}{\partial \mathbf{x}}\boldsymbol{\Sigma}_{xx}\left(\frac{\partial \mathbf{l}}{\partial \mathbf{x}}\right)^\mathsf{T} \\ &= \mathsf{S}_x\boldsymbol{\Sigma}_{yy}\mathsf{S}_x^\mathsf{T} + \mathsf{S}_y\boldsymbol{\Sigma}_{xx}\mathsf{S}_y^\mathsf{T} \end{aligned}$$

which with (12) yields

$$\boldsymbol{\Sigma}_{ll} = \sigma_x^2 \begin{pmatrix} 2 & 0 & -x_1 - y_1 \\ 0 & 2 & -x_2 - y_2 \\ -x_1 - y_1 & -x_2 - y_2 & {y_1}^2 + {x_1}^2 + {y_2}^2 + {x_2}^2 \end{pmatrix}$$

assuming stochastic independence of $\mathbf{x}$ and $\mathbf{y}$.

Point at inifinity: The point $\mathbf{X}_\infty$ has covariance matrix

$$\boldsymbol{\Sigma}_{X_\infty X_\infty} = 2\sigma_{x^d}^2 \begin{pmatrix} \underset{2\times2}{I} & \underset{2\times2}{\mathbf{0}} \\ \underset{2\times2}{\mathbf{0}} & \underset{2\times2}{\mathbf{0}} \end{pmatrix}.$$

4.3 Jacobians for Optimal Estimation

We now provide the Jacobians $\mathsf{A}_i(\boldsymbol{y}_i)$ and $\mathsf{B}_i(\boldsymbol{y}_i, \mathbf{p})$ for the individual constraints.

4.3.1 Observed Vertical Lines We have the vector $\boldsymbol{y}_i^\mathsf{T} = (\boldsymbol{l}_i^\mathsf{T}, \boldsymbol{x}_i^{d\mathsf{T}})$ of the measured entities, relevant for vertical lines. Therefore, we obtain from (7)

$$\mathsf{A}_i(\boldsymbol{y}_i) = \mathsf{A}_i(\mathbf{l}_i, \boldsymbol{x}_i^d) = \frac{\partial \mathbf{e}_i}{\partial \mathbf{p}} = \begin{pmatrix} \mathbf{l}_i^\mathsf{T}\otimes\mathbf{U}_i^\mathsf{T} \\ \mathbf{l}_i^\mathsf{T}\otimes\mathbf{V}_i^\mathsf{T} \end{pmatrix}.$$

The other Jacobian can be obtained by

$$\begin{aligned} \mathsf{B}_i(\boldsymbol{y}_i, \mathbf{p}) = \mathsf{B}_i(\mathbf{l}_i, \boldsymbol{x}_i^d, \mathbf{p}) &= \frac{\partial \mathbf{e}_i}{\partial \boldsymbol{y}_i} = \left(\frac{\partial \mathbf{e}_i}{\partial \boldsymbol{l}_i}\ ,\ \frac{\partial \mathbf{e}_i}{\partial \mathbf{U}_i}\frac{\partial \mathbf{U}_i}{\partial \boldsymbol{x}_i^d}\right) \\ &= \begin{pmatrix} \mathbf{U}_i^\mathsf{T}\mathsf{P}^\mathsf{T} & \mathbf{l}_i^\mathsf{T}\mathsf{P}(I_{2\underset{2\times2}{\mathbf{0}}})^\mathsf{T} \\ \mathbf{V}_i^\mathsf{T}\mathsf{P}^\mathsf{T} & \mathbf{l}_i^\mathsf{T}\mathsf{P}(I_{2\underset{2\times2}{\mathbf{0}}})^\mathsf{T} \end{pmatrix} \end{aligned}$$

using the Jacobian $\partial\mathbf{U}_i/\partial\boldsymbol{x}_i^d$ from (6). Note that we here only consider $\mathbf{l}_i$ and the coordinates $X = x_i^d$ and $Y = y_i^d$ on the drawing as uncertain. The heights of the 3D points, Z_1 and Z_2, may be fixed arbitrarily and, therefore, are treated as deterministic values.

4.3.2 Observed Horizontal Lines We have the vector $\boldsymbol{y}_j^\mathsf{T} = (\mathbf{l}_j^\mathsf{T}, \mathbf{X}_{\infty j}^\mathsf{T})$ as uncertain observations. We, therefore, obtain from (8)

$$\mathsf{A}_j(\boldsymbol{y}_j) = \mathsf{A}_j(\mathbf{l}_j, \mathbf{X}_{\infty j}) = \frac{\partial \mathbf{e}_j}{\partial \mathbf{p}} = (\mathbf{l}_j \otimes \mathbf{X}_{\infty j})^\mathsf{T},$$

and

$$\mathsf{B}_j(\boldsymbol{y}_j, \mathbf{p}) = \mathsf{B}_j(\mathbf{l}'_j, \mathbf{X}_{\infty j}, \mathbf{p}) = \frac{\partial \mathbf{e}_j}{\partial \boldsymbol{y}_j} = \left(\mathbf{X}_{\infty j}^\mathsf{T}\mathsf{P}^\mathsf{T} \ , \ \mathbf{l}_j^\mathsf{T}\mathsf{P}\right).$$

4.3.3 Observed Points We have the vector $\boldsymbol{y}_k^\mathsf{T} = (\mathbf{x}_k,^\mathsf{T}, \mathbf{X}_k^\mathsf{T})$ as uncertain measurements. Therefore, we obtain from (9)

$$\mathsf{A}_k(\boldsymbol{y}_k) = \mathsf{A}_k(\mathbf{x}_k, \mathbf{X}_k) = \frac{\partial \mathbf{e}_k}{\partial \mathbf{p}} = \overline{\mathsf{S}}_{x_k} \otimes \mathbf{X}_k^\mathsf{T}.$$

The other can be obtained by

$$\mathbf{B}_k(\mathbf{x}_k, \mathbf{X}_k, \mathbf{p}) = \frac{\partial \mathbf{e}_k}{\partial \boldsymbol{y}_k} = \left(-\overline{\mathsf{S}}_{PX_k} \ , \ \overline{\mathsf{S}}_{x_k}\mathsf{P}\right). \quad (13)$$

This holds for arbitrary point correspondences.

4.4 Optimization function

Using the assumed covaraince matrices for the observed entities and the derived Jacobians, we can derive the covariance matrices $\boldsymbol{\Sigma}_{e_ie_i}$, $\boldsymbol{\Sigma}_{e_je_j}$, and $\boldsymbol{\Sigma}_{e_ke_k}$ of the constraints' residuals $\boldsymbol{e}_i$, $\boldsymbol{e}_j$ and $\boldsymbol{e}_k$ as given in (3).

We now have all components at our disposal necessary to estimate P. Thus, we need to solve iteratively

$$\begin{aligned}
&\Big[\sum_i \mathsf{A}_i^\mathsf{T}(\widehat{\boldsymbol{y}}_i^{(\nu)}) \left(\widehat{\boldsymbol{\Sigma}}_{e_ie_i}^{(\nu)}\right)^+ \mathsf{A}_i(\boldsymbol{y}_i) \\
&+\sum_j \mathsf{A}_j^\mathsf{T}(\widehat{\boldsymbol{y}}_j^{(\nu)}) \left(\widehat{\boldsymbol{\Sigma}}_{e_je_j}^{(\nu)}\right)^+ \mathsf{A}_j(\boldsymbol{y}_j) \\
&+\sum_k \mathsf{A}_k^\mathsf{T}(\widehat{\boldsymbol{y}}_k^{(\nu)}) \left(\widehat{\boldsymbol{\Sigma}}_{e_ke_k}^{(\nu)}\right)^+ \mathsf{A}_k(\boldsymbol{y}_k) \quad \Big]\widehat{\boldsymbol{p}}^{(\nu+1)} = \lambda\widehat{\boldsymbol{p}}^{(\nu+1)},
\end{aligned} \quad (14)$$

where $\boldsymbol{p}$ contains the unknown parameters of P. The iteration can be initialized by results from a direct estimation of $\boldsymbol{p}$ as in (11), which is equivalent to setting $\boldsymbol{\Sigma}_{e_ie_i} = \boldsymbol{\Sigma}_{e_je_j} = \boldsymbol{\Sigma}_{e_ke_k} = \mathsf{I}$. The fitted observations $\widehat{\boldsymbol{y}}_i$, $\widehat{\boldsymbol{y}}_j$, and $\widehat{\boldsymbol{y}}_k$, are initialized by the actual observations in the image and the drawing.

5 EXPERIMENTS

In this section we describe the experiments conducted to evaluate the characteristics of the presented method. We first show the correctness of error prediction by testing the method on synthetic data. Then, the practical usability of self-diagnosis tools is demonstrated by showing results of application to images and a drawing of a high voltage switch gear.

5.1 Synthetic Data

In order to guarantee realistic test conditions we use geometric dimensions which come very close to those given in the real data experiments in the following section. We use a drawing of 400 $\times$ 400 pixels, the camera is placed off the drawing at approximately $(252, -222, 108)$[pixel] and has a viewing angle of approximately 50 degrees. We generate uniform distributed points on and above the ground, vertical lines, and horizontal lines. We may then project all features into the camera, using a known realistic projection matrix, to have an ideal set of observations.

As a first indication for correctness of the estimated covariance of P, $\widehat{\boldsymbol{\Sigma}}_{\widehat{p}\widehat{p}}$, we compare the estimated back-projection error of a 3D point with its true covariance under well defined noise conditions. Therefore, we estimate the projection matrix from randomly perturbed observations, and back-project a perturbed 3D point using the estimated projection. Using a perturbation of $\sigma_d = 0.5$[pixel] for observations from the drawing, and $\sigma_i = 1.2$[pixel] for image features, we repeat the estimation of P $n = 5000$ times using 10 vertical lines, 10 horizontal lines, and 10 points. The resulting Gaussian distributed point cloud is depicted in Fig. 3. The experiment shows that the predicted 90% confidence region from (5) and the empirically obtained one are very close. Using the same setup, we now consider the Mahalanobis distance between the estimated projection parameters and the true ones $\tilde{\mathbf{p}}$:

$$||\widehat{\mathbf{p}} - \tilde{\mathbf{p}}||^2_{\Sigma_{\beta\beta}} = (\widehat{\mathbf{p}} - \tilde{\mathbf{p}})^\mathsf{T}\widehat{\boldsymbol{\Sigma}}^+_{\widehat{p}\widehat{p}}(\widehat{\mathbf{p}} - \tilde{\mathbf{p}})$$

If the model is correct, the Mahalanobis distance is χ^2_{11} distributed, since the projection matrix has 11 degrees of freedom. The histogram in Fig. 3 shows the distribution of the Mahalanobis distance based on 1000 experiments. The comparison with the analytically plotted χ^2_{11} distribution shows good conformity.

5.2 Real Data

A large field of application for the presented pose estimation method is as-built reconstruction in industrial environments. Therefore, we chose images of a high voltage switch gear and the associated top view drawing to demonstrate the feasibility of the algorithm. The drawing has a dimension of 1220$\times$820 pixel. Its scale [pixel]:[m] is 1:0.012, i. e. 1 pixel is equivalent to 12mm in the real world. The images have a size of 1530$\times$ 1018 pixel. We select feature points and lines in the image and in the drawing with a standard deviation of 1 pixel and 0.5 pixel, respectively. We assume that the selection process itself on the drawing does not have any uncertainty, since drawings are in general vectorized documents. However, drawings reflect the as-built situation only up to a certain accuracy. In our case, the drawing depicts the real world installation up to approx. $\pm$ 6mm, which results in the mentioned standard deviation.

As observations in the image we select 6 vertical lines, 5 horizontal lines, 3 points on the ground and 2 above. Once

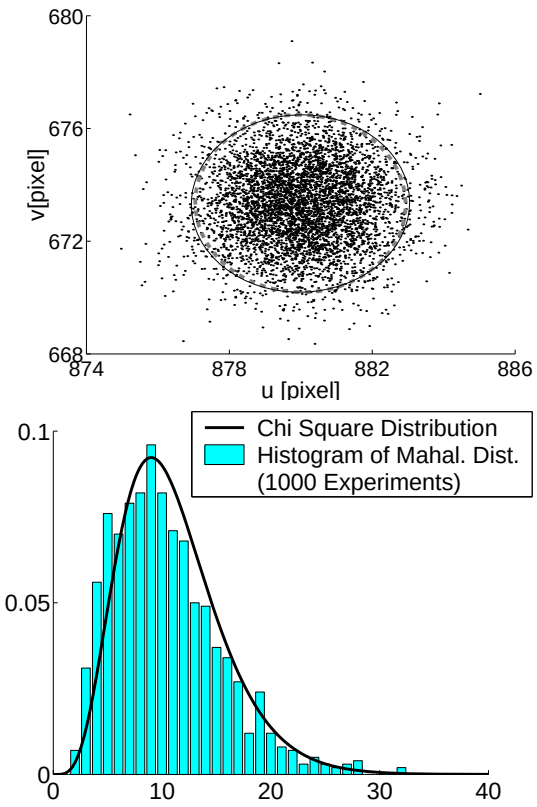

Figure 3: **Top:** Back-projection of a 3D point using estimated projection matrices. The projection is estimated based on randomly perturbed observations in the image and in the drawing. 5000 samples are taken to obtain the empirical 90% confidence region (solid), which is very close to the predicted one (dashed) from (5). **Bottom**: The Mahalanobis distance of the estimated projection matrix and the true projection matrix is χ^2_{11} distributed. The histogram represents the distribution obtained by 1000 experiments on perturbed data. The curve indicates the true χ^2_{11} distribution. The mean Mahalanobis distance is 10.87, compared to 11, which is statistically not different.

the projection matrix $\widehat{\mathsf{P}}$ and the associated covariance matrix $\widehat{\mathbf{\Sigma}}_{\widehat{p}\widehat{p}}$ are estimated, we can obtain the uncertainty of any 3D point projected in the image. To visualize this, we select 3D points on the ground plane and points above of each, in heights of 100, 200 and 300 pixels. The uncertainties of their projections are shown in Fig. 4. Note that the ellipses are plotted magnified by a factor of 10 for better visibility.

Self-diagnosis can be bound to the estimated variance factor $\widehat{\sigma}_0^2$ (cf. (4)). Here we obtain $\widehat{\sigma}_0^2 = 1.5$, thus $\widehat{\sigma}_0 = 1.25$. Assuming the constraints actually hold, we can conclude that the standard deviation of the measurements are worse by a factor $\widehat{\sigma}_0^2 = 1.25$, which practically confirms the assumed standard deviation. An indicator for characterizing the accuracy of the estimated projection matrix is the uncertainty of the camera's center of projection $\widehat{\mathbf{Z}}$ (cf. A.1). Computing the standard deviations for the camera center of the image shown in Fig. 4, we obtain standard deviations of $\sigma_X \approx 11$mm, $\sigma_Y \approx 37$mm, and $\sigma_Z \approx 10$mm, where the indices X, Y, and Z indicate the coordinate axes of the

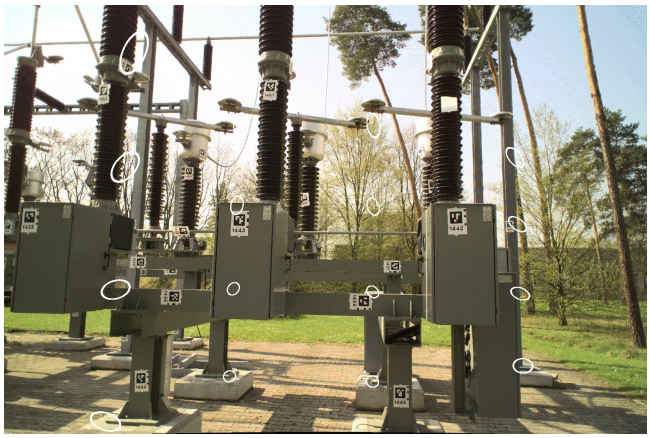

Figure 4: Uncertainty of projected 3D points. The ellipses show the covariances of 3D points on the ground, and points above of each in heights of 100, 200, and 300 pixels. The ellipses are plotted magnified by a factor of 10 for better visibility. In this case, the ground plane is not located on the actual ground, but at the top edge of the concrete foundations.

drawing. In this case, the camera's principal axis is roughly parallel to the Y-axis of the drawing. Due to the fact that the intrinsic parameters, in particular the principle distance c, is estimated, this appears to be a reasonable result and is fully acceptable in many applications. For augmented reality applications, it is very useful to superimpose the image of an installation and the associated drawing. Fig. 5 shows such an overlay resulting from application of our method to another scene. All information contained in the drawing is directly connected to the respective components in the real world. At the same time, blending drawings and images can be used as a visual accuracy check of the projection matrix.

6 CONCLUSION

We presented a new method for single image orientation using scene constraints in conjunction with a drawing of the scene. We use vertical lines, located in the drawing or the map, the direction of horizontal lines, which may also be contours of cylinders, and points. Besides a direct solution, we provided a statistically optimal estimation procedure for the projection matrix P. This allows us to verify the results statistically, in particular to test whether the assumed noise model is correctly chosen. We applied the method to synthetic data to validate the correctness of error prediction. Statistical tests on experiments using real data showed that good results can be obtained using a comparatively small number of constraints. The availability of this type of statistical tools for self-diagnosis and for charaterizing the performance increases the usefulness of image analysis considerably.

A APPENDIX

A.1 Uncertainty of the Camera Center

Let the projection matrix be partitioned in the classical way

$$\mathsf{P} = (\mathsf{H} \mid -\mathsf{H}\boldsymbol{Z}) = (\mathsf{H} \mid \mathbf{h}).$$

Figure 5: Superimposition of an image and the associated drawing of the scene. Again, the ground plane is not located on the actual ground, but at the top edge of the concrete foundations.

Then the projection center $\boldsymbol{Z}$ fulfills

$$\mathbf{h} + \mathsf{H}\,\boldsymbol{Z} = \mathbf{0} \qquad \text{or explicitely} \qquad \mathbf{Z} = -\mathsf{H}^{-1}\,\mathbf{h}.$$

From total differentiation $d\mathbf{h} + d\mathsf{H}\,\boldsymbol{Z} + \mathsf{H}\,d\boldsymbol{Z} = \mathbf{0}$ we obtain with $\mathbf{Z} = (\boldsymbol{Z}^\mathsf{T}, 1)^\mathsf{T}$ and $\mathrm{vec}(\mathsf{P})^\mathsf{T} = ((\mathrm{vec}\mathsf{H})^\mathsf{T}, \mathbf{h}^\mathsf{T})$ the differential

$$\begin{aligned}
d\boldsymbol{Z} &= -\mathsf{H}^{-1}\,d\mathsf{H}\,\boldsymbol{Z} - \mathsf{H}^{-1}\,d\mathbf{h} \\
&= -(\mathsf{H}^{-1} \otimes \boldsymbol{Z}^\mathsf{T})\,\mathrm{vec}(d\mathsf{H}) - \mathsf{H}^{-1}\,d\mathbf{h} \\
&= -\mathsf{H}^{-1}\,(I_3 \otimes \boldsymbol{Z}^\mathsf{T} \,|\, I_3 \otimes 1)\,\mathrm{vec}(d\mathsf{P}) = -\mathsf{H}^{-1}\,(I_3 \otimes \mathbf{Z}^\mathsf{T})\,\mathrm{vec}(d\mathsf{P}) \\
&= -\mathsf{H}^{-1}\,(\mathbf{Z}^\mathsf{T} \otimes I_3)\,\mathrm{vec}(d\mathsf{P}^\mathsf{T}) = -\mathsf{H}^{-1}\,(\mathbf{Z}^\mathsf{T} \otimes I_3)\,d\mathbf{p}
\end{aligned}$$

from which the covariance matrix

$$\boldsymbol{\Sigma}_{\widehat{Z}\widehat{Z}} = \mathsf{H}^{-1}\,(\mathbf{Z}^\mathsf{T} \otimes I_3)\,\boldsymbol{\Sigma}_{\widehat{p}\widehat{p}}\,(\mathbf{Z} \otimes I_3)\,\mathsf{H}^{-\mathsf{T}}$$

follows.

REFERENCES

Bondyfalat, D., Papadopoulo, T. and Mourrain, B., 2001. Using scene constraints during the calibration procedure. In: ICCV01, pp. II: 124–130.

Caprile, B. and Torre, V., 1990. Using vanishing points for camera calibration. IJCV 4(2), pp. 127–140.

Faugeras, O., 1993. Three-Dimensional Computer Vision - A Geometric Viewpoint. Artificial Intelligence, The MIT Press.

Faugeras, O. and Luong, Q. T., 2001. The Geometry of Multiple Images. MIT Press.

Förstner, W., 2001. Algebraic projective geometry and direct optimal estimation of geometric entities. In: S. Scherer (ed.), Computer Vision, Computer Graphics and Photogrammetry – a Common Viewpoint, ÖAGM/AAPR, Österreichische Computer Gesellschaft.

Hartley, R. and Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University Press.

Kanatani, K., 1996. Geometric Computation for Machine Vision. Oxford Science Publication.

Klette, R., Schlüns, K. and Koschan, A., 1998. Computer Vision : Three-Dimensional Data from Images. Springer.

Matei, B. and Meer, P., 1997. A general method for errors-in-variables problems in computer vision. In: CVPR.

van den Heuvel, F., 1999. Estimation of interior orientation parameters from constraints on line measurements in a single image. International Archives of Photogrammetry and Remote Sensing 32,Part 5W11, pp. 81–88.

Wang, L. and Tsai, W., 1991. Camera calibration by vanishing lines for 3-d computer vision. PAMI 13(4), pp. 370–376.

Youcai, H. and Haralick, R., 1999. Testing camera calibration with constraints. PhEngRS 65(3), pp. 249–258.