Computer Vision and Photogrammetry – Mutual Questions: Geometry, Statistics and Cognition*

Wolfgang Förstner
Institut für Photogrammetrie, Universität Bonn
Nussallee 15, D-53115 Bonn, e-mail: wf@ipb.uni-bonn.de

Preface: The emerging interaction between Computer Vision and Photogrammetry certainly is well in the flavor of Kennert Torlegard's professional life: Not only his PhD thesis dealt with un-calibrated cameras, not only was one of his main interests close range photogrammetry with all its various applications, no, he also was active in bringing the researchers of both fields together. This paper on one side collects experiences of the dialog between Computer Vision and Photogrammetry. On the other side it gives an example, closely related to Kennert Torlegards PhD thesis [Torlegard, 1967] of a type of analysis hopefully useful for researchers from both fields, illuminating the common fields geometry and statistics and the possibilities of mutual exchange, and finally reflects on the recent developments in the area of cognitive vision and their relation to aerial image interpretation.

1 Motivation

This paper gives a personal view on the dialog of Computer Vision and Photogrammetry experienced in nearly two decades. It wants to emphasize the necessity of a still closer interaction between both fields and tries to illuminate classical pitfalls in discussions resulting from the still too large number of mutual pre-justices.

Interestingly, in 1993 two CV researchers, J. Mundy and R. Hartley, on a SPIE Conference tackled the same topic [Mundy and Hartley, 1993], also motivated by the strong interest to increase mutual cooperation. This paper differs in several ways, it is written nearly 10 years later, it contains a photogrammetric view, however, tries to address both communities.

2 FAQ

At least since the 1983-workshop Photogrammetry meets Pattern Recognition in Graz, the two communities Photogrammetry and Computer Vision are in a dialog. Starting a communication between representatives of two cultures always

*Paper presented at the International Symposium *Photogrammetry meets Geoinformatics*, 17. April 2002, dedicated to Prof. Kennert Torlegard.

is difficult as the words stand for different notions. Views have a tradition but consciousness of the roots always remains limited.

Age, size and homogeneity of the two communities are different. In spite of nearly two decades the relationship between the communities still appears to be remote. Few researchers in both fields have dived into the other field and tried to grasp the flavor, many know the basic interests of the other group. But they have different context in their research, there is a lack of transparent references, especially concerning the photogrammetric literature, the main reference for researchers in Computer Vision still being the Manual of Photogrammetry of the ASPRS from 1980. These differences of the two communities make pre-justices inevitable.

We want to illuminate a number of pre-justices between the two communities coming up the last years. They are put into questions posed explicitely or implicitly during the last years, often appearing between the lines in discussions and can be seen as a virtual panel discussion between representatives of Photogrammetry (PH) and Computer Vision researchers (CV). Questions and answers want to illuminate views and reasons, at least aiming at a mutual understanding for the current situation.identification of required system behavior.

Questions to Photogrammetrists Being polite, we start with questions posed by our (virtual) guests.

CV: *Why do you only aim at highest precision?*

PH: We do not only aim at highest precision.

In general, we aim at the precision just necessary for the task in order to be economical. To achieve this we need to know the best possible precision achievable with a certain system. As a by-product, this allows us to fully exploit the precision inherent in a system. With good systems we therefore are able to solve practical problems where highest precision is necessary to be competitive with other mensuration techniques. This especially holds for close range applications where we use targeted points or for mapping, where the costs for acquiring the images are a significant percentage of the total costs.

CV: *Why do you only use calibrated cameras?*

PH: This is motivated by the main application of photogrammetry, namely mapping where using un-calibrated cameras[1] would lead to severe deficiencies: Already orientation of a single un-calibrated camera in a nearly flat terrain is extremely unstable. Moreover, calibrating a camera, though certainly additional effort, is possible, as no real time evaluation of the images is necessary and we have access on the planning stage of the data acquisition.

But be aware, in close range applications the handling of un-calibrated cameras is well known, the direct linear transformation (DLT) proposed more than 30 years ago is identical to the mapping $\mathbf{x}' = \mathsf{P}\mathbf{X}$ from object to image space using the projection matrix P. Also the relations to the parameters of the exterior and interior orientation have been developed, admittedly not as transparent by the Computer Vision community in the last decade (cf. below).

With the increasing use of CCD- and video-cameras the role of un-calibrated cameras in photogrammetry might increase, if the range of applications appears

[1] The notion *un-calibrated camera* is used differently in the both fields: whereas in Computer Vision and un-calibrated camera (usually) still is straight line preserving, in Photogrammetry an un-calibrated camera just is a camera where the calibration is unknown.

to be promising enough ...

CV: *Why are you not interested in automatically obtaining approximate values?*

PH: This is only partially true. Obtaining approximate values always has been a problem. The situation is different in the two main application areas:

In aerial triangulation, the basis for mapping, obtaining approximate values in the 70's has been solved by simultaneous adjustment of all images, not bundles, in a planimetric block adjustment. Since about a decade the position of the projection centers are directly measured obtained by GPS, angles are close to zero or measured by INS.

In close range photogrammetry, there exist some few program packages which fully automatically determine approximate values for the orientation parameters. However, one must admit: they start at the measured image points, thus presume the correspondence problem to be solved. As this application area appears small efforts to automate this important process are limited.

CV: *Why are you not interested in direct minimal solutions?*

PH: The interest in direct solutions highly depends on whom you ask: Researchers in the evaluation of aerial image assume good approximate values for the orientation procedures to be available (cf. above). Only those researchers who work in the area of automatic orientation use direct solutions for orientation procedures. The DLT, the estimation of the essential matrix and the direct estimation of the spatial similarity transformation are sufficient for most applications.

Admittedly the interest in direct solutions in other areas, e. g. in the area of feature extraction or 3D-reconstruction is not broad, yet, at least it is not one of the commonly accepted conceptual issues in building automatic algorithms.

CV: *Why are you only interested in geometry?*

PH: This is a classical misunderstanding. It results from the two views on Photogrammetry, also present in the area of surveying or geodesy, where we usually are part of:

Photogrammetry in a (narrow) technical sense deals with the geometry of images. This distinguishes Photogrammetry from Remote Sensing, which in a (narrow) technical sense deals with the radiometry of images.

Photogrammetry in a (broad) engineering sense deals with the information extraction from images. Mapping, the largest market segment within Photogrammetry, deals with the recovery of semantic information. Of course, when filling Geoinformation Systems the geometric component plays an essential role, however, position only is one, admittedly important, attribute out of many.

CV: *Why do the new developments on geometry not seem useful or necessary for Photogrammetry*

PH: The impression is correct, but it does not reflect the real situation especially in the long run.

First, orientation procedures have been developed in the last three decades and made very sophisticated and, what is most important, transferred to application, where they are used in daily work.

Second, Photogrammetric research during the last decade was dominated certainly by three topics: (1) automatic orientation of aerial images including automatic matching techniques and geodetic measurements (GPS, INS), (2) airborne laser ranging, becoming a heavy competitor to photogrammetric stereo analysis and (3) automatic and semiautomatic cartographic feature extraction especially road and building extraction.

Some Photogrammetrists grasped the new developments in projective geometry, some curricula now include parts of these developments. However, the process of integrating the new orientation procedures into basic photogrammetry needs a careful evaluation and has just started.

CV: *Why do you claim to have done everything before and CV researchers do not know old literature?*

PH: This statement certainly over emphasizes, is only partly true, may be the result of being uncertain about the real innovations in CV and may be overcome.

The essential matrix for relative orientation has been published by Thomson in 1962, trilinear relations between coordinates in an image triplet have been given by Mikhail in 1962/63, critical surfaces for single and stereo image orientation are known since a long time.

But, we have to admit, the framework developed in CV is much richer and general, especially including un-calibrated cameras, but also the cases of calibrated cameras. Also we have to admit, that the transparency of our publications is not high enough to grasp the essentials of basic geometry, which makes access by the CV community very difficult. We also know, that in the meantime CV researchers have studied the old photogrammetric literature written by mathematicians, even found errors, and reestablished these old solutions in a new context.

CV: *Why do you say, CV researchers have little knowledge of statistical analysis*

PH: Again, this statement is overemphasizing and certainly dates back a few years.

Many solutions, e. g. the algebraic minimization techniques in the direct solutions, do not at all take the uncertainty of the measurements into account. Performance characteristics using redundant systems and covariances has not been standard. On the other hand, classical results from adjustment theory have been re-derived.

But we have to admit, that the developments in the last few years have changed the situation. However, we have the impression that the exchange in the area of statistics still could be intensified.

CV: *Why do so few Photogrammetrists join CV conferences?*

PH: CV as a topic is only one part of Photogrammetry, as we are embedded in the larger field of Mapping and GIS.

On the other hand from the area of image analysis, which certainly is of great interest to us in general, only a small percentage actually touches our current problems, which are more of the systems engineering type ore are related to procedures really applicable in practice. E. g. the matching techniques required for automatic aerial triangulation are available or need to be adapted to the specific boundary conditions, this adaption appears not to be interesting for the CV community. Another example is the area of building extraction: The topic already is very specific, from the CV community it appears as a special application area not a basic research area, this is why these developments are discussed on specific workshops, such as Stockholm 1995 [Torlegard and Gülch, 1995] and ASCONA 1996 and 1999 [Baltsavias et al., 2001]. As already mentioned, the main stream research in photogrammetry was only partly overlapping the stream in CV.

Questions by Photogrammetrists Now, let us take the chance to ask our (virtual) guests. Let us start with topics closest to photogrammetry.

PH: *Why do you like to use un-calibrated cameras?*

CV: We start with the most general situation: Assume you get an image or two, what can you tell about object space? In many applications we have no information about the camera or we are not able to use calibrated cameras, e. g. when we want to exploit the zoom capability.

Moreover, assuming the cameras to be un-calibrated, allows to exploit the full potential of projective geometry. This gives us access to direct solutions, such as the DLT. Moreover, we easily can integrate straight lines or cylinders into the analysis.

On the other hand, we do not like un-calibrated cameras per se, therefore we have developed schemes for self calibration, which however are setup differently, starting with un-calibrated cameras obtaining projective results and then upgrading them to Euclidean results. This eases automation of getting approximate values. We do not assume the images to be in a special setup, as e. g. in your mapping applications.

PH: *Why do you not care about precision?*

CV: Having heard your answer about precision, we can argue identically: Our application defines the precision. Obviously our domains of applications are very different: E. g. in case you want to perform obstacle avoidance with a moving robot you do not need the position of the obstacle better than 5 % of the distance. Therefore simple, suboptimal solutions often are better than needed.

But we have to admit, optimal solutions have the advantage of being easily characterized. You certainly know, that since a few years we use bundle adjustment, minimizing the re-projection error, as a reference or as final step in all orientation and calibration tasks.

PH: *Why are minimal solutions of such high interest?*

CV: All robust techniques being not maximum-likelihood type estimators, such as RANSAC, minimizing least median squares or clustering, require a direct solution of the unknown parameters from a set of given observations or constraints. The high algorithmic complexity of these algorithms force to find solutions with the minimum possible number of constraints. The break down point of these algorithms, though theoretically below 50 %, practically can be mor than 70 or 80 %, making these estimates extremely powerful.

PH: *Why do you say, Photogrammetrists have no knowledge of projective geometry?*

CV: The basic impression is that the orientation problems are solved within the photogrammetric community, starting with known approximate values and solving large non-linear systems of equations iteratively.

Perhaps, it is not necessary for you to use projective geometry at all. But, communication about orientation procedures appears to be difficult, if not done on the basis of projective geometry as documented in the two new books on multiple view geometry by Hartley/Zisserman and by Faugeras/Luong.

PH: *Why are you only interested in theoretical results?*

CV: This is an interesting question. In the 80's CV was characterized by the exploration of image analysis tools with only little theory below. This lead to a strong reaction, condensed in the two papers *CV-theory – the lack thereof* by Haralick and *I have seen yor demo – so what?* by Price and influenced the developments in the 90's, especially on the main conferences, ICCV and ECCV. In the mean time papers without a broad empirical basis have much less chance in getting accepted at these conferences, slowing down theoretical progress in an adequate manner. Still, the interest is in innovative developments also in theory, but covering more than a single subtask, and look for the integration aspect which is the prerequisite for improving applicability.

PH: *Why do you think, in the mean time there is no interest anymore in statistical analysis?*

CV: This is only partly true.

Error analysis is understood in principle, characterizing results with their distribution is a classical technique, performance characteristics of algorithms is clear from its setup.

But, all other techniques, such as Bayes'ian nets, Monte Carlo techniques, support vector machines or the whole area of learning, e. g. in data mining and image retrieval are still not fully exploited.

PH: *Why is CV not any more interested in image understanding?*

CV: This question is as interesting as the one on theory in CV and certainly closely related.

As you know, image understanding research was very active in the 80's. Around 1990 one realized the big gap between low level vision and feature extraction on one side and image understanding or vision as part of artificial intelligence. The lack of theory for the basic tasks of feature extraction, segmentation, tracking, stereo, form representation etc. motivated the type of research with very much concentrated on very specific vision subtasks. Unification of boundary detection segmentation and stereo including texture analysis, the strong developments in geometry, the successful appearance based techniques for recognition, the recent break through's in tracking using statistical sampling techniques, e. g. with condensation algorithms, and, always parallel, the purposive vision for navigation, tracking of grasping, lead to the insight to restart research in systems integration and high level vision. The EU started an initiative on cognitive AI-based vision both by supporting research and by establishing a network for promoting the field. This might be an interesting possibility for you Photogrammetrists to rethink mapping from images for filling GIS data bases (cf. below).

PH: *Why do so few CV researchers join photogrammetric conferences?*

CV: We seem to have the same situation as you have: Photogrammetry in the narrow sense of calibration, orientation and geometric reconstruction as a topic is only a small part of CV. Most CV researchers see themselves (1) as engineers in image science trying to make automatic vision feasible, (2) as natural scientists which try to model the human system by studying the effects of both or (3) a mixture of both, trying to learn from nature for increasing insight into what vision could achieve and to establish new and hopefully better solutions. The photogrammetric community appears to be very much focused on the analysis of aerial images, what you called close range photogrammetry is appears not to be present in our field.

Some of us have been to your conferences and realized and very much appreciated the intensive experimental part in all scientific papers. Perhaps you could make us curious on how you could help us in solving interesting vision problems, e. g. in preparing standard reference data sets for image analysis as a basis for performance characteristics which appears as a common interest, and present it at some of our conferences in order to increase motivation to join yours.

3 Exchanging Experience in Geometry and Statistics

There are classical common topics in image analysis which could be the basis for a mutual exchange of experience. Two of them are geometry and statistics. The examples given now want to demonstrate this using camera orientation as an example.

3.1 Import of Experience from CV: Geometric Modeling Cameras

The collinearity and the coplanarity equations are the basic equations for geometric computations in photogrammetry. They may be represented in several ways. We want to show that the developments in Computer Vision in the meantime documented in two textbooks [Hartley and Zisserman, 2000, Faugeras and Luong, 2001] are of great advantage already for representing the geometry of one and two views.

The collinearity equation The classical collinearity equation with five parameters for the interior orientation in the Computer Vision literature is written with homogeneous coordinates (drawn in upright letters) $\mathbf{x}' = (u', v', w')^{\mathsf{T}} = w'(x', y', 1)^{\mathsf{T}}$ $\mathbf{X} = (U, V, W, T)^{\mathsf{T}} = T(X, Y, Z, 1)^{\mathsf{T}}$ and the 3×4 matrix $\mathsf{P} = (P_{ij})$ as

$$\mathbf{x}' = \mathsf{P}\mathbf{X} \qquad \text{or} \qquad \mathbf{x}' \times \mathsf{P}\mathbf{X} = \mathsf{S}_{x'}\mathsf{P}\mathbf{X} = \mathbf{0} \tag{1}$$

the second expression being a constraint between object and image points, where the skew-symmetric matrix $\mathsf{S}_{x'}$ induces the cross product. We will see, that such constraints can be formulated for other entities easily and used as basis for a bundle adjustment using various types of geometric features.

Now, the projection matrix can be explicitely expressed in terms of the 11 parameters of the exterior and interior orientation, as

$$\mathsf{P} = \mathsf{K}R(I| - \boldsymbol{X}_O) \qquad \text{with} \qquad \mathsf{K} = k_{33}K = k_{33}\begin{pmatrix} c & cs & x'_H \\ 0 & (1+m)c & y'_H \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

where the upper triangular matrix K is the calibration matrix, having an arbitrary scale via k_{33}, and containing principle distance, principle point, shear and scale difference. It is easy, to derive the 11 parameters from a given projection matrix $\mathsf{P} = (\mathsf{A}|\mathbf{a})$, partitioned into a left 3×3 matrix and a 3-vector:

$$\boldsymbol{X}_O = -\mathsf{A}^{-1}\mathbf{a} \qquad \mathsf{A} = \mathsf{K}R \qquad K = \mathsf{K}/k_{33} \tag{3}$$

where the second expression can be used to derive K and R from A via a QR-decomposition.

Obviously, the matrix notation allows to write the collinearity equation in a very compact form, to relate the elements of the projection matrix of the DLT immediately to the elements of the exterior and interior and exterior orientation, and thus establish *full equivalence between the several representations.*

Observe, we did not make any assumptions on whether the interior orientation of the camera is known or not. Thus the relations hold for calibrated and uncalibrated cameras. Thus *the projective representation is no argument against using it for calibrated cameras.*

The coplanarity equation The coplanarity equation usually is derived for so-called reduced image coordinates, i. e. ${}^k\boldsymbol{x}'$ and ${}^k\boldsymbol{x}''$. However, given the exterior and the interior orientation of two cameras no direct expression for the coplanarity equation seems to be known in the photogrammetric literature. This might be very useful, in case one wants to automatically derive 3D-coordinates from images after an aerial triangulation, as then the projection matrices are known and one needs the coplanarity equations for establishing the epipolar geometry.

Given two images with projections $\mathbf{x}' = \mathsf{P}_1\mathbf{X} = \mathsf{K}_1 R_1(I| - \boldsymbol{X}_{O1})\mathbf{X}$ and $\mathbf{x}'' = \mathsf{P}_2\mathbf{X} = \mathsf{K}_2 R_2(I| - \boldsymbol{X}_{O2})\mathbf{X}$ the directions ${}^0\mathbf{x}' = R_1^{\mathsf{T}}\mathsf{K}_1^{-1}\mathbf{x}'$, ${}^0\mathbf{x}'' = R_2^{\mathsf{T}}\mathsf{K}_2^{-1}\mathbf{x}''$ and the basis $\mathbf{b} = \boldsymbol{X}_{O2} - \boldsymbol{X}_{O1}$ we immediately can write $w = [\ {}^0\mathbf{x}' \quad {}^0\mathbf{b} \quad {}^0\mathbf{x}''] = \mathbf{x}'^{\mathsf{T}}\mathsf{K}_1^{-\mathsf{T}} R_1 \mathsf{S}_b R_2^{\mathsf{T}} \mathsf{K}_2^{-1}\mathbf{x}'' = \mathbf{x}'^{\mathsf{T}}\mathsf{F}\mathbf{x}'' \overset{!}{=} 0$. Again this is a bilinear form with the central matrix

$$\mathsf{F} = \mathsf{K}_1^{-\mathsf{T}} R_1 \mathsf{S}_b R_2^{\mathsf{T}} \mathsf{K}_2^{-1} \tag{4}$$

being the so-called *fundamental matrix.* It makes the relation to the given parameters of the exterior and interior orientation explicit.

For the normal case with $R_1 = R_2 = I$ and $\mathsf{K}_1 = \mathsf{K}_2 = \mathrm{Diag}(c, c, 1)$ and the basis $\mathbf{b} = (B_X, 0, 0)^{\mathsf{T}}$ the coplanarity reduces to $w = (y'' - y')B_X/c \overset{!}{=} 0$ which is the classical requirement, the y-parallaxes should vanish. In a similar manner linearization of w easily allows to derive linearized observation equations for least squares adjustment.

The representations developed in the area of Computer Vision only partially have been proposed earlier in Photogrammetry. They obviously have not found their way into the university education in Photogrammetry as a standard.

One might speculate why. One reason certainly is inertia, as most orientation procedures in practical applications can do without the new developments, much more it would be additional work to re-implement existing procedures without direct advantages; thus there appears no real need. The new representations partially are more abstract and do not relate to classical mathematical education, especially the use of homogeneous coordinates is not part of the standard curriculum of a Photogrammetrist.

On the other hand only linear algebra is required, the connections to computer graphics require the knowledge of projective geometry anyway, and – what is most important – the insight into the internal structure of the geometry of the imaging process eases the understanding. Moreover, extending Point-Photogrammetry to Line-Photogrammetry, a process having started more than a decade ago [Mulawa, 1989, Zielinski, 1993], enormously profits from using the new representation schemes (cf. below). Finally, the direct access of the computer to digital imagery, especially CCD-images and video sequences allows to increase the application area of Close Range Photogrammetry from, say, below 5 % to much more than 10 % starting to overlap with the application areas of Computer Vision; there pre-calibration of cameras may not be directly feasible.

3.2 Export of Experience to CV: Statistics of Camera Orientation

We now want to analyze the effect of using calibrated versus un-calibrated cameras for the orientation of a single. We refer to the normal case and use a three dimensional control point field in the form of a flat cube, as a two dimensional control point field does not allow to determine the orientation of un-calibrated cameras.

This type of *analyzing the theoretical precision* of estimates is common in Geodesy and Photogrammetry and up to now is not standard in Computer Vision. The case discussed here is very close to the one analyzed by K. Torlegard in his PhD thesis [Torlegard, 1967].

The situation is sketched in fig. 1. The control point field consists of 8 control points, observed from a camera at height Z_O above the center of gravity. The control points arranged in two squares with half width $D = wZ$ and half height $H = hD = hwZ$, thus $h = 1$ would represent an cube and $w = \tan\alpha/2$ measures the viewing angle. The principle distance is c.

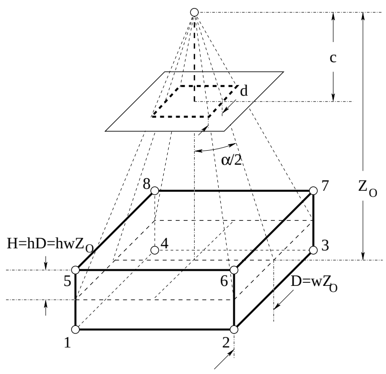

Figure 1: Setup for the orientation of one image with a three dimensional control point field. The 8 corners of the flat square cube is observed at distance Z_O from the projection center.

Assuming independent equally distributed measured image coordinates of all 8 points we obtain the standard deviations for the 6 parameters of the exterior orientation in dependency of the geometric parameters c, w, h, Z_O and σ_0.

We distinguish the spatial resection (SRS) with a calibrated camera from the direct linear transformation (DLT) for a un-calibrated camera. The expressions for a calibrated camera using a spatial resection (SRS) are lengthy and not given here. In case of $h = 0$ they specialize to the well known relations

$$\sigma_{X_O}^{SRS} = \sigma_{Y_O}^{SRS} = \frac{\sqrt{2}}{4}\frac{Z_O}{c}\sqrt{1 + \frac{1}{\sin^4\frac{\alpha}{2}}}\,\sigma_0 \qquad \sigma_{Z_O}^{SRS} = \frac{1}{4}\frac{Z_O}{d}\,\sigma_0$$

deviating a factor $\sqrt{2}$ from the classical case with four control points.

The standard deviations for the case of an un-calibrated camera with 11 unknown parameters using the DLT and the 3D-control point structure are

$$\sigma_{X_O}^{DLT} = \sigma_{Y_O}^{DLT} = \frac{\sqrt{2}}{4}\frac{Z_O}{c}\frac{|1-(hw)^2|}{hw}\sqrt{\frac{1+10(hw)^2+(hw)^4}{1+6(hw)^2+(hw)^4}}\,\sigma_0$$

and

$$\sigma_{Z_O}^{DLT} = \frac{1}{4}\frac{Z_O}{c\,d}\frac{|1-(hw)^2|}{hw}\sqrt{1+(hw)^2}\,\sigma_0$$

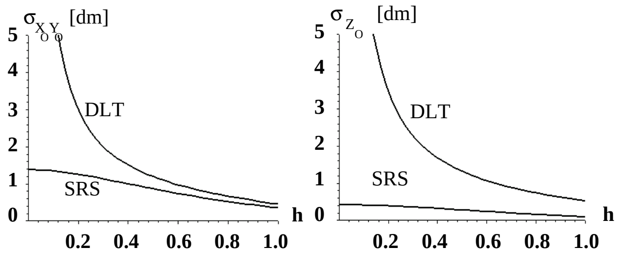

Figure 2: Theoretical accuracy $\sigma_{X_O Y_O}$ and σ_{Z_O} in dm as a function of the relative height h of the reference object, wide angle camera, scale 1:10000, $\sigma_0 = 10$ μm

In fig. 2 the theoretical precision for the position of the projection center is given as a function of the relative height h of the reference object. We assumed $w = 0.6$ corresponding to the case of an aerial image, $\sigma_0 Z_O/c = 1$ corresponding to assuming $\sigma_0 = 10$ μm, $c = 0.15$ m and $ZO = 1500$ m and measuring the precision in dm.

(1) The theoretical precision of the position of the camera for an un-calibrated camera (DLT) is much worse than for calibrated cameras (SRS) . As to be expected, no determination is possible for a flat object for an un-calibrated camera.

(2) The precision of the depthZ_O with a calibrated camera (SRS) and a flat object ($h = 0$) can be achieved with an un-calibrated camera only if the object has full depth ($h = 1$).

(3) The precision of the pose across the viewing direction achievable with a calibrated camera for a flat object can be achieved with an un-calibrated camera only if the reference object has approximately half the depth than the width ($h = 1/2$).

This type of analysis, being standard in Photogrammetry and Geodesy, may be easily performed in order to get a first insight into the structure of the geometric setup[2]. Of course this type of analysis is not as far reaching as the algebraic analysis which is able to immediately cover the whole space of parameter values and therefore should performed if feasible.

3.3 Fusion of Experience: Far Beyond Points for Bundle Adjustment

We now want to give an example how to fuse the two fields of experience, projective geometry and statistics especially adjustment theory. As an example we use the bundle adjustment in an industrial environment with – possibly marked – points, straight lines and cylinders, cf. the pioneering photogrammetric work in [Mulawa, 1989, Zielinski, 1993].

The Gauß-Helmert Model The Gauß-Helmert model in adjustment theory has the following general form

[2] An experienced Photogrammetrist would expect exactly this weakness of the relative orientation.

$$\boldsymbol{g}(\boldsymbol{l}+\boldsymbol{e},\boldsymbol{x}) = \mathbf{0} \qquad \boldsymbol{h}(\boldsymbol{x}) = 0 \qquad \Omega = \boldsymbol{e}^{\mathsf{T}}\boldsymbol{\Sigma}_{ll}^{-1}\boldsymbol{e} \tag{5}$$

thus containing general nonlinear constraints between observed values, collected in the vector $\boldsymbol{l}$ and the unknown parameters $\boldsymbol{x}$. Minimizing Ω leads to the classical maximum likelihood solution in case the errors $\boldsymbol{e}$ in the observations are normally distributed.

The model is the most general case, and only in case one can solve the functions $\boldsymbol{g}(\boldsymbol{l}+\boldsymbol{e},\boldsymbol{x})$ for the observations can be reduced to the Gauß-Markoff-model (with constraints). Nearly, all constraints naturally can be expressed in the structure of the Gauß-Helmert model, as the following example shows [Appel et al., 2002].

Constraints for Points, Lines and Cylinders For *points* we just use (1):

$$\mathsf{S}_{x'}\mathsf{P}(\boldsymbol{p})\mathbf{X} = \mathbf{0} \tag{6}$$

Now we have observe points $\mathbf{x}'$, the parameters $\boldsymbol{p}$ of the projection matrix and the object points are unknown, and we possibly have constraints on the norm of P: Thus we have constraints as in the Gauß-Helmert model (5). In case of calibrated cameras, the parameters $\boldsymbol{p}$ may the 6 parameters of the exterior orientation and no constraint on P is required.

3D lines best are represented with a 6-vector in Plücker coordinates. Assume we have an unknown 3D line $\mathbf{L}$ and observe straight line segments with starting and end point $\mathbf{x}$ and $\mathbf{y}$ we need to require that the start and end points sit on the image $\mathbf{l}'$ of the 3D-line. The image $\mathbf{l}'$ of a 3D-line $\mathbf{L}$ reads as [Förstner, 2000, Faugeras and Luong, 2001]

$$\mathbf{l}' = \mathsf{Q}(\boldsymbol{p})D\mathbf{L} \qquad \text{with} \qquad \mathsf{Q}(\boldsymbol{p}) = \begin{pmatrix} (\mathbf{B}\cap\mathbf{C})^{\mathsf{T}} \\ (\mathbf{C}\cap\mathbf{A})^{\mathsf{T}} \\ (\mathbf{A}\cap\mathbf{B})^{\mathsf{T}} \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} \mathbf{0} & I \\ I & \mathbf{0} \end{pmatrix}$$

The 3×6 projection matrix $\mathsf{Q}(\boldsymbol{p})$ for lines depends on the rows of the projection matrix for points $\mathsf{P}^{\mathsf{T}}(\boldsymbol{p}) = (\mathbf{A},\mathbf{B},\mathbf{C})$ and $\mathbf{A}\cap\mathbf{B}$ is the intersection line of the two planes $\mathbf{A}$ and $\mathbf{B}$. Thus we have $\mathbf{x}'^{\mathsf{T}}\mathbf{l}' = 0$ and $\mathbf{y}'^{\mathsf{T}}\mathbf{l}' = 0$ or

$$\mathbf{x}'^{\mathsf{T}}\mathsf{Q}(\boldsymbol{p})D\mathbf{L} = 0 \qquad \mathbf{y}'^{\mathsf{T}}\mathsf{Q}(\boldsymbol{p})D\mathbf{L} = 0 \qquad |\mathbf{L}| = 1 \qquad \mathbf{L}^{\mathsf{T}}D\mathbf{L} = 0 \tag{7}$$

In addition we have to constraints for the 3D-line being of the form $\boldsymbol{h}(\boldsymbol{x}) = 0$, the first taking the homogeneity of the line vector into account and the second being the Plücker constraint. Thus we again achieve the structure of the Gauß-Helmert model. Also here the projection matrices may be only depend on the parameters of the exterior orientation.

Cylinders can be represented by their axis $\mathbf{M}$ and their radius r (cf. the early work [Mulawa, 1989]. We now assume to observe a point $\mathbf{x}'$ in an image which sits on the apparent contour of the cylinder. Then the spatial distance $d(\mathbf{L}',\mathbf{M}) = r$ of the projection ray $\mathbf{L}'(\boldsymbol{p}) = \mathsf{Q}(\boldsymbol{p})^{\mathsf{T}}\mathbf{x}'$ from the cylinder axis $\mathbf{M}$ should equal r. Thus we obtain the constraint

$$d^2(\mathbf{L}'(\boldsymbol{p}),\mathbf{M}) = \frac{|\mathbf{L}'^{\mathsf{T}}(\boldsymbol{p})D\mathbf{M}|^2}{|\boldsymbol{L}'_h(\boldsymbol{p}) \times \boldsymbol{M}_h|^2} = r^2 \qquad |\mathbf{M}| = 1 \quad \text{and} \quad \mathbf{M}^{\mathsf{T}}D\mathbf{M} = 0 \tag{8}$$

where the line parameters a split into two 3-vectors, e. g. $\mathbf{M}^{\mathsf{T}} = (\boldsymbol{M}_h^{\mathsf{T}}, \boldsymbol{M}_0^{\mathsf{T}})$. Taking the square avoids a decision on the sign of the bilinear form $\mathbf{L}'^{\mathsf{T}}(\boldsymbol{p})D\mathbf{M}$. Also here the two constraints for the 3D-line $\mathbf{M}$ need to be used. Implementation of such a bundle adjustment can be done easily. Results using calibrated cameras can be found in [Appel et al., 2002].

4 Cognition

What is to come in the future? All photogrammetric problems which do not require some type of image interpretation have been successfully addressed in the last two decades and lead to working software. However, one must admit, that the success lies in the intelligent interaction between computer and operator, clearly separating tasks, and allowing final editing. In Computer Vision the low- and mid-level processes have reached a quite high standard. However, the goal of scene understanding, in the sense of reconstructing the scene to an extend which is able to explain the image data, has not achieved yet. In both areas earlier attempts in the late 80' and the early 90' to automate image interpretation got stuck because of the gap between the error prone low-level processes including feature extraction and the high-level AI-type processes which could not handle the imperfectness of the available data and had severe problems in modeling real complex situations.

There seem to be two trends in Computer Vision and Photogrammetry which might be the basis for a fruitful cooperation: cognitive AI-based vision and mapping on demand.

4.1 Cognitive AI-Based Vision

The EC has identified the area of cognitive AI-based vision as research area for the next years. It supports projects and has started a network, called ECVision, with the following goals (taken from the proposal written by David Vernon): A.: *Research Planning* - identify key challenges, problems, and system functionalities, B.: *Education and Training* - identify and develop courses, curricula, texts, material, and delivery mechanisms, C.: *Information Dissemination* - promote the visibility and profile of cognitive vision at conferences and in journals, D.: *Industrial Liaison* - identify application drivers and highlight any successes, promote research trials, addressing all types of industries: games, entertainment, white goods manufacturers (e.g. vigilant appliances), construction (e. g. smart buildings), medicine (e.g. aids for the disabled), etc.

The characteristics of cognitive AI-based visions are (again taken from the proposal):

- *Knowledge representation of events and structures* exhibiting some form of object or category invariance with respect to events and/or vision system behavior,
- *Learning* including learning to see and learning to do,
- *Reasoning* about events and about structures,
- *Recognition and categorization:* Cognitive vision systems would ideally have categorization capability, being a difficult issue because objects of the same category can have completely different visual appearances, and
- *Goal specification*, i. e. identification of required system behavior.

4.2 Mapping on Demand

Providing the data for Geoinformation Systems, short mapping, as one of the basic tasks of Photogrammetry, has drastically changed its flavor in the last decade:

data have to be acquired nearly in real time, the group of users of image data products has increased dramatically, as a consequence the product requirements have reached a diversity which seems to make the concept of a base map obsolete, in case one takes the prefixed legend as part of the map.

On the other hand all attempts in automatic image interpretation, road or building extraction, appear to show the same type of diversity, due to use of different data sources, different image analysis procedures and also different specifications for the final results; the cause for this diversity mainly is the difficulty in achieving some reasonable results in an exploration type of research.

It appears to be a challenge to fuse the trends: Use cognitive AI-based vision to help solving the difficult problems in mapping on demand.

- Representing knowledge about topographic objects, their appearance and analysis processes
- Learning models and analysis processes
- Reasoning about the interaction between computer and operator
- Recognition of generic topographic objects
- Specification of user demands

Photogrammetry should take a driving role in order to establish close interaction with researchers in CV and AI in order to achieve the next step in photogrammetric image analysis.

References

[Appel et al., 2002] Appel, M., Förstner, W. and Navab, N., 2002. Cylinders, lines and points for the orientation and reconstruction of industrial sites. Technical report, Inst. Photogrammetry, Bonn University.

[Baltsavias et al., 2001] Baltsavias, E., Grün, A. and Gool, L. V. (eds), 2001. Automatic Extraction of Man-Made Objects from Aerial and Space Images. Balkema.

[Faugeras and Luong, 2001] Faugeras, O. and Luong, Q., 2001. The Geometry of Multiple Images. MIT-Press.

[Förstner, 2000] Förstner, W., 2000. New Orientation Procedures. In: Int. Archives for Photogrammetry and Remote Sensing, Vol. XXXII.

[Hartley and Zisserman, 2000] Hartley, R. and Zisserman, A., 2000. Multiple View geometry in Computer Vision. Cambridge Press.

[Mulawa, 1989] Mulawa, D., 1989. Estimation and Photogrammetric Treatment of Linear Features. PhD thesis, Purdue, West Lafayette, Indiana, USA.

[Mundy and Hartley, 1993] Mundy, J. and Hartley, R., 1993. The relationship between photogrammetry and computer vision. http://www.syseng.anu.edu.au/ hartley/My-Papers.html, SPIE.

[Torlegard, 1967] Torlegard, K., 1967. On the Determination of Interior Orientation of close-up Cameras under Operational Conditions using Three-dimensional Test Objects. Stockholm.

[Torlegard and Gülch, 1995] Torlegard, K. and Gülch, E. (eds), 1995. The Role of Models in Automated Scene Analysis.

[Zielinski, 1993] Zielinski, R., 1993. Object with Digital Line Photogrammetry. PhD thesis, KTH, Stockholm.