# Automatic Orientation and Recognition in Highly Structured Scenes

W. Förstner and E. Gülch

Institute for Photogrammetry, University of Bonn
D-53115 Bonn, Germany

## ABSTRACT

The paper discusses the impact of scene and assessment models for videometry. Full automation of calibration and orientation procedures appears to be as necessary for enlarging the field of applications as the use of explicit geometric and semantic scene knowledge. The focus on achieving highest possible accuracy needs to be embedded into a broader context of scene analysis. Examples demonstrate the feasibility of tools from Computer Vision for image metrology.

**Keywords:** Computer Vision and Photogrammetry, camera calibration, camera orientation, 3D-modeling, object- and image models, segmentation, semantic scene knowledge

## 1. INTRODUCTION

Optical 3D-mensuration techniques play an increasing role in industrial, medical, or environmental applications. The direct access of high performance computers to the acquired measurements, intensities in CCD-cameras or distances in range finders, provide the potential for full automation of all processes, especially camera calibration and orientation, surface reconstruction, metrical inspection, deformation analysis, or monitoring quality production processes and especially tracking, navigation etc.

The acceptance of optical 3D-metrology, however, is still limited. The number of reasons is high, and often is related to problems of specification and underestimating the systems aspects, e. g when not taking into account that: *the following equation is always true: Vision-system does not equal PC + Frame-grabber + Camera + Software.*[1]

An as severe reason for vision systems not to be accepted to the desired degree is the user unfriendliness of many systems: camera orientation and calibration is still an art, objects need to be targeted, change of illumination conditions make systems fail without being able to tell the user why, focus on high precision prevents realizing potential applications with accuracy requirements below 1 : 1000, surface mensuration techniques only gradually yield vector data, object identification still is not really approached. There are of course counter examples to this list.[2–6] But they appear to be prototypical or not yet built as parts into systems.

The difficulty with relying solely on image data and large number of applications where the objects are comparably small, lead to an increase interest in the use of range data. As they immediately yield 3D-data, the difficult image matching problem is circumvented. However, range finders also yield iconic data, i. e. data which are unstructured, in contrast to vector or vectorized data. Therefore nearly all above mentioned problems – except the hard illumination problem – still remain: calibration, orientation, fusion of 3D-data, interpretation.

Whereas the photogrammetric community seems to concentrate on the development of high precision sensors and corresponding calibration and orientation procedures, the Computer Vision community is interested in automating processes, especially orientation processes, and exploiting specific or generic knowledge of the scene. Progress in integrating procedures started on one side in the matching area, where tools from the Computer Vision community were taken, adapted, or conceptually rebuilt. The proper statistical analysis of the procedures and the high precision requirements which had to be fulfilled lead to successfully working systems.[7–11] Also the flavor of the knowledge based systems for image analysis have severely influenced research in photogrammetry.[12] The restricted tools for

Other author information: (Send correspondence to W. F.)
W. F.: Email: wf@ipb.uni-bonn.de; Telephone: +49-228-73-2714; Fax: +49-228-73-2712;
E. G.: Email: ebs@ipb.uni-bonn.de; Telephone: +49-228-73-2904; Fax: +49-228-73-2712;
**Videometrics V, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III, SPIE-publication 3174-01, San Diego 1997, SA.**

fully automatically determining orientation parameters from point measurements lead to strong developments in the area of projective geometry in the Computer Vision community.[13–15]

This paper is motivated by the strong belief that an intimate exchange of ideas between both communities will increase progress in both areas and broaden the potential of vision metrology.

From the many problems area we selected three issues to be discussed in this context:

1. Automation of orientation and calibration procedures.

   Non-specialists who would like to work with a vision system do not want to take care of a preprocessing step, which only yields *nuisance* parameters – with respect to their application, they do not want to pay a specialist and would like to use the system in a flexible manner, like a measuring rod. It is an open question whether fully automatic orientation and calibration systems can be designed in a manner that they work under general conditions. Therefore any increase of automation within semiautomatic systems would help.

   Accuracy of the original measurements among other effects decides on the overall achievable accuracy. We therefore discuss some pitfalls in automatic high precision pointing. Fully autonomous calibration and orientation of cameras seems to be achievable. This refers to the complete chain from the digital image data to the final determination of the camera parameters. Projective geometry eases both the transparent derivation of orientation procedures as well as the actual computation. Automation requires self diagnosis. An example shows the successful use of classical statistical concepts in geodesy applied to a complex orientation procedure.

2. Recognition for 3D-mensuration.

   Object recognition has not really been a topic in metrology, as the goal of metrology is to obtain metric not thematic information. However, it is well accepted that the quality of models may directly influence the quality of the resulting measurements. The model usually is given by the systems designer and only occasionally may be influenced by the user, e. g. when choosing a calibration model. In highly structured scenes the reconstruction of the geometry of the scene depends on the assumed structure, which in interactive systems is given by the human operator, who e. g. indicates the image measurements are related to a cylinder in space.

   Geometric information about the scene is needed when deriving a geometric description of the scene in vector format. But this information also increases the precision of the achieved measurements. Moreover, even thematic information helps making more precise measurements. We will discuss to what extent semantic models for the scene are necessary for achieving reliable 3D-descriptions for virtual reality applications or rapid prototyping.

3. 3D-Models for automatic 3D-reconstruction.

   Metrology usually is applied in highly structured domains, excluding trees, ash-pits or the like. Therefore the development of adequate models appears to be feasible. Such models in an first instance relate to the 3D-structure of the scene. For automatic reconstruction, however, the appearance in the intensity or range images need to be explicitly modeled. Moreover, any image analysis step leads to a modified, often symbolic description of the images or the scene, which therefore also need to be modeled in order to be usable by the automatic system.

   We therefore discuss possible models suitable for automatic but also semiautomatic image analysis. These models may be used in all steps why we specifically discuss their use for sensor orientation and 3D-reconstruction. The discussion will reveal the necessity to develop models which reflect the different aggregation layers of composite objects in order to be able to automatically identify object parts in the image data.

## 2. MODELING

Recovering a description of the scene from sensor data can be seen as inverting the imaging process (cf. fig. 1). The object, being described by name, class, form and pose is imaged using some sensor, being described by its interior and exterior orientation, leading to some iconic description, e. g. a raster image or a triangulated surface. The inversion of the imaging process obviously only is possible if, besides knowledge about the imaging process, some preknowledge about the scene is available. Whereas image metrology up to now concentrated on determining sensor orientation and pose and form of objects, it appears that additional information such as class of object or object parts is required.

**Figure 1.** Forward and backward modeling. Objects are described by name, class, form and pose which possibly together with camera parameters have to be recovered partially or completely by image analysis


## 2.1. Object, Sensor, Image and Analysis Models

Therefore all steps in the data acquisition chain need to be modeled. We distinguish object models, sensor models, image models and analysis models[16]:

- Object models describe the general structure of the scene as far it is essential for solving the specific application. E. g. in deformation monitoring representing the object as a point cloud may be sufficient, whereas in deformation analysis the object needs to be described as consisting of mutually disjunct rigid or non-rigid bodies, with surface physical properties which allow observation by a fixed installed camera system.

- Sensor models describe the specific properties of the sensing process thus include the lighting model. Classical Photogrammetry restricted modeling to the geometric aspects, whereas Remote Sensing focussed on the physical aspects. Automatic videometry requires both, as already measuring image points depends on an adequate model for the geometric and radiometric properties of the used digital image. Sensor modeling, however, shows its importance when integrating sensors of different type within one system and sensor planning, as well known for aerial triangulation and industrial applications of Photogrammetry, play a central role. The early work of Ikeuchi[17] on formalizing sensor modeling is motivated by the need to automatically optimize sensor selection, sensor orientation and image analysis for object recognition.

- Image models represent the appearance of the object in the sensed data. Image models therefore formally can be derived from the object and the sensor model. Most image analysis procedures start with some, more or less explicit, assumptions about the appearance of the object in the image, even identify the image of the object with the object itself. The appearance based methods in Computer Vision, which heavily trigger current research, are motivated by the difficulty to derive the image model in a formal manner and the observation that many task in object recognition can be solved without explicit reference to the 3D-structure. In image metrology this approach is limited, as precise and complete geometric information cannot be recovered without explicit reference to the object and the sensor model.

- Analysis models represent the strategies for deriving a description of the object from the sensed data. In classical Photogrammetry this contains e. g. point identification and measurement, determination of approximate values for orientation parameters and object points, bundle adjustment and final evaluation. Recovering a vector representation of a surface is a segmentation problem which requires decisions on how to define similarities of neighboring surface patches, methods to detect surface discontinuities and tools for handling occluded areas, inconsistencies of the topology or disturbing objects, not fulfilling the assumed surface model. Obviously, vision metrology here can learn and actually learns from the research in Computer Vision.

When realizing these models two aspects need special attention:

- Local and global constraints:

  Occlusions, illumination effects, and image noise prevent the observations and therefore recognition and reconstruction of complete objects in single steps. Therefore the granularity of the modeling needs to be high enough in order to identify substructures of the objects in the image data. Consequently the object and the image model needs to describe local and global constraints. E. g. dominant directions or regularities as parallelism or symmetry, which – using the sensor model – can be transfered to constraints in the image, allowing to identify substructures which then may be used to derive partial information about the scene, e. g. camera orientation or class of object part, and trigger the following steps in the analysis. Examples will be given in the following sections.

- Geometric, physical and semantic models:

  In order to fulfill the requirement of the user geometric models are not sufficient. Data acquisition for spatial information systems, rapid prototyping and virtual reality applications may require more than a set of 3d-points or a triangulated surface possibly with texture mapping, namely identification of objects and their relations. Already in deformation analysis the user is not interested in the 3d-coordinates of the points but in the segmentation of the object into a set of moving rigid or deforming bodies, which reveals this problem to be a complex data interpretation problem.[18,19] It will be shown which role recognition of geometric structures and semantic knowledge may play for 3d-modeling.

We will discuss object models relevant for image metrology in more detail in order to demonstrate the role of these models during automatic image analysis.

## 2.2. Objects Models for Image Metrology

Object models for metrology have three essential parts, geometry, reflectance and semantics. It has shown to be useful to distinguish different types of models, depending on the type of techniques used for reconstruction or recognition:

- *Specific models* describe the object except for its pose. All parameters describing the form, the reflectance properties or the type of the objects are known. Examples are targets, CAD-models or known Digital Surface or Elevation Models. Parameter estimation techniques are used for recovering the pose. The small number of pose parameters allows to use robust techniques with high break down point[20] or invariants for recognition.[15]

- *Parametric models* describe the object up to its pose and a fixed set of shape or reflectance parameters. Examples are targets of given form but unknown size, e. g. circles, primitives in CAD-systems, e. g. boxes or cylinders. But also the iconic description of a free form surface using a specified grid with unknown depth or height values is a parametric model. Also here reconstruction leads to a parameter estimation problem, however, due to the possibly large number of the parameters only maximum-likelihood-type robust estimators may be used and invariants refer to local geometric structures, e. g. rectangles between neighboring edges or just the topology of the object, when described in a vector format, e. g. specifying the degrees of nodes or the number of sides of polygonal faces.

- *Generic models* contain unknown cardinal numbers or structures in addition to continuous parameters for specifying pose or shape. We may distinguish two cases:

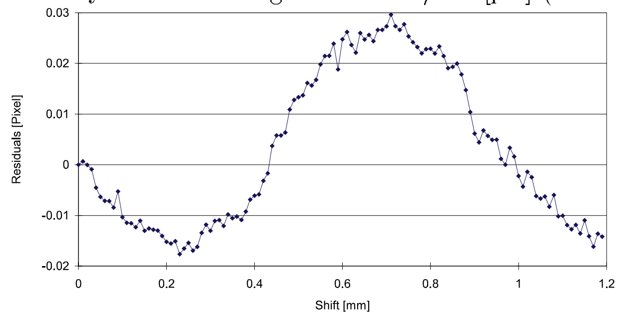
  - Models which, besides continuous parameters, have one or several *cardinal numbers as unknown.*

    Classical examples are polynomials of unknown degree or polygons, e. g. when identifying the optimal distortion model during calibration, with unknown numbers of edges, e. g. when identifying the ground plan of a building. In both cases object reconstruction needs to identify the unknown cardinal number of parameters and the actual values of these parameters. The optimal reconstruction leads to a combined classification and estimation problem, which may be solved using the principle of Minimum Description Length or, in case a priori knowledge on the number of parameters is available, by Bayesian classification. An example is given for the interpretation of 3D-edges of man made objects (cf. section 4.2).

  - Models which do not specify the structure of the object in detail.

    Classical examples are polyhedra, representing an indoor environment, or complex objects consisting of parts which may be represented as a CSG-tree (constructive solid geometry) of parametric primitives. In this most general situation the reconstruction problem not only needs to identify the number of parts, e. g. number of faces of the polyhedron or the number of object parts, but also their mutual relations. These relations not only may refer to the geometry or the physical properties but in many cases to the semantics of the parts and their relations.

**Figure 2.** Cross correlation residuals showing a periodic bias with a period of one pixel. The maximum bias is appr. 0.03 [*pel*]. The random errors obviously are in the range of below 1/200 [*pel*] (Courtesy of Casott and Prenting[27]).




## 3. AUTOMATIC CALIBRATION AND ORIENTATION

We discuss means for reliably achieving high precision and automation in calibration and orientation procedures, namely the role of bias for accuracy, the role of projective geometry fro obtaining approximate values, the role of generic scene knowledge for orientation and the potential of tools for self diagnosis in automatic image analysis.

### 3.1. Automatic mensuration: The role of bias

Todays mensuration systems show a consistently high accuracy. Targets can be measured with standard deviations between 0.01 and 0.05 [*pel*][21,22] and thus confirm simulations[23] and theoretical predictions[24] on target location. Depending on the size of the CCD-array, the quality of the used camera model and the mensuration design relative accuracies up to more than 1 : 100 000 are achievable. As the theoretical predictions suggest standard deviations down to 0.01 [*pel*] to be reachable increase of precision may be achieved by a more careful analysis of the target location techniques.

Bias due to linearization usually is not to be expected in high precision applicatios, as the bias increases with the relative precision of the measurements.[25] In image analysis this, however, does not hold any more. E. g., the relative accuracy of finding edge pixels at the boundary of a circle is in the range of 1/10 to 1/1000. Effects of bias have been studied occasionally.[26]

Casott/Prenting[27] report a periodic bias during target location by cross correlation as shown in figure 2. The empirical bias was obtained by moving the target with high precision and measuring its position in the digital camera. The average bias of appr. 0.015 [*pel*] is about 3 or 4 times larger than the random errors being in the range below 1/200 [*pel*]. The maximum bias has been found to depend on the thickness of the cross shaped target: thinner crosses lead to higher bias, which contradicts theoretical expectations.

The classical way to achieve sub pixel accuracy during cross correlation is to fit a parabola through the point of maximal correlation and its neighbors and to determine the maximum of this parabola. In case of thin targets the cross correlation function is sharply peaked suggesting high precision to be achievable. However, the approximation of the correlation function by a parabola leads to a bias in the position.

The bias can be theoretically predicted. Assume the correlation function is centered at 0, is bell shaped with width $\sigma$. Assume the pixel distance is $\Delta x = 1$ and the grid is shifted by $a \in [\frac{-1}{2}, \frac{1}{2}]$ The optimal point of the correlation function $c(x)$ is determined by

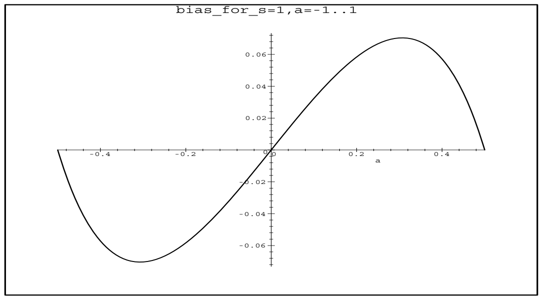
$$\widehat{x} = x^o - \frac{c'(a)}{c''(a)} = x^o - \frac{\frac{1}{2}[c(a+1) - c(a-1)]}{c(a+1) - 2c(a) + c(a-1)} \tag{1}$$

where $x^o$ is the approximate integer position, here $x^o = 0$. If $c$ is a parabola $c = kx^2$ then (1) gives the peak of the parabola, independent on the choice of $a$.

Now assume the bell shaped correlation function to be

$$c(x) = \frac{1}{1 + \frac{1}{2}\frac{x^2}{\sigma^2}} \tag{2}$$

**Figure 3.** Bias of cross correlation for a correlation function with width $\sigma = 1$. The maximum bias appr. is 0.06 [*pel*]




then the bias induced by the approximation of $c(x)$ by a parabola through the points $a-1$, 0 and $a+1$ is given by

$$b(\widehat{x}) = a\frac{1-4a^2}{2s^2-3a^2+1} \tag{3}$$

It has its maximum at $\frac{1}{12}\sqrt{54+144s^2-6\sqrt{33+336s^2+576s^4}}$.

The maximum bias can reasonably well be approximated by

$$b_{\max} = \frac{1}{3+12s^2} \tag{4}$$

Obviously the bias increases inversely with the square of the width of the correlation function. Figure 3 shows the bias for a correlation function with $\sigma = 1\,[pel]$. It corresponds to the empirical findings shown in fig. 2. The maximum bias approximately is 0.06 [*pel*] showing the severe influence of the width. For a smaller value of $\sigma = 1/2\,[pel]$, thus for a very narrow correlation function, the bias is nearly 1/5 [*pel*], for a larger value of $\sigma = 2\,[pel]$, however, the bias is still 0.02 [*pel*]. The bias effect can be taken into account and used to obtain an unbiased estimate for the position.

### 3.2. Automatic calibration: The role of projective geometry

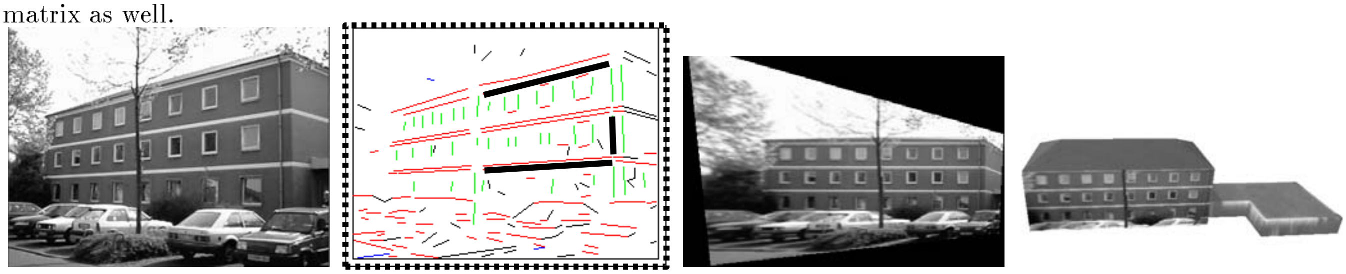
Camera calibration, i. e. the determination of the intrinsic parameters or the parameters of the interior orientation, is indispensable for achieving accurate measurements. The number of calibration procedures is large. However, in case one is interested in calibrating one's own camera, such calibration procedures appear not to be available or not easy to handle. As intrinsic parameters usually are nuisance parameters, camera calibration should be easy, fast, reliable and as automatic as possible.

There are several modes in which camera calibration could be performed: Using a test field with targeted points leads to most accurate results. Only approximate values for the coordinates need to be available for the targets due to the self calibration procedure which determines all parameters simultaneously. The evaluation of the data may be performed nearly automatically, except for the acquisition of the image data.[2] In case the camera would be mounted on a robot full automation would be possible, even more, the system could decide whether calibration is needed and initiate the data acquisition, leading to an autonomous calibration scheme.

The hard part is the automatic numbering of the image points using some matching procedure. Here approximate values for the 3D-coordinates are necessary for most systems. Some systems rely on specially coded targets. Planar test fields with planar targets allow to use projective geometry to advantage, as the homogeneity of the projection equations lead to linear systems to be solved when determining transformation parameters. This holds for both, target recognition, for reconstruction and for determining approximate values.[2,28]

In case of calibration in non targeted but structured domain calibration has to rely on well identifiable but natural points. Thus the only requirement for the domain is that enough distinct points are detectable, which can be

**Figure 4.** The orientation of a single image of our institute (a, left) was performed using automatically extracted straight line segments (b, middle left), used for rectification of the facade (c, middle right) and merged with texture form an aerial image to obtain a photo realistic rendering (d, right). Identifying the three bold lines (b, middle left) as related to the object's coordinate axes – in an interactive mode – allows to automatically derive the rotation matrix as well.



identified as homologous point in the images. Then fully automatic procedures are able to perform the matching and to determine accurate enough approximate values[3] for a subsequent self calibrating bundle adjustment. Also here, the matching procedure heavily relies on knowledge from projective geometry in order to prune the number of matching candidates, e. g. by exploiting the epipolar constraint between three views using the trilinear tensor.[13,29,30] Of course the accuracy of such a calibration will not be as high as with targeted points, but sufficient for the reconstruction of the scene in concern. The argumentation relies on the long experience with self calibrating bundle adjustment in aerial triangulation which could be transferred to the automatic aerial triangulation techniques.[31–33]

### 3.3. Automatic orientation: The role of scene knowledge

Camera orientation requires some scene knowledge, usually some control points, for relating the camera coordinate systems to the object coordinate system. Identification of control information cannot easily be automated as control information has to be described in a general manner, which due to the different application areas is not possible. This contrasts to the determination of the relative position of the cameras which can be automated under certain conditions (cf. above).

From the remaining 7 degrees of freedom for translation, rotation and the rotation can be derived automatically from images alone in case the scene reveals regularities, specifically if part of it are of *Legoland* nature, i. e. show three dominant mutually orthogonal directions. Then with a vanishing point analysis the rotation matrix can be derived. The idea of deriving camera parameters is older than photography, going back to Lambert in the 18th century,[34] and has attracted many researchers especially in Computer Vision.

The idea of deriving the rotation matrix from vanishing points uses the fact, that the direction of sets of lines in 3D is parallel to the direction from the projection center to the corresponding vanishing point and the three orthogonal vectors to the three vanishing points form the columns of the rotation matrix, assuming the intrinsic parameters of the camera are known.

An example is given in fig. 4a, which was taken with an digital amateur camera, Kodak DC25, which has shown to have very low distortions. Fig. 4b shows the automatically extracted straight line segments.[35,36] Most of them point towards the vanishing point to the left, many towards the zenith, some to the vanishing point at the right side. An automatic clustering[37] yielded the rotation matrix which was used to rectify the image, shown in fig. 4c. No interaction is necessary for these steps.[38] Using the approximate position of the camera from a GPS-receiver would allow to match the image to the 3D-model of the building. Fig. 4d shows a textured view, where the front texture was taken from a terrestrial image and the roof textures from an aerial image. The 3D-model of the building was acquired interactively from an aerial image pair.

In case no approximate camera position is known only an interactive solution seems to be feasible. It however requires only the precise measurement of two points for specifying the rectangle in the rectified image and the indication on which surface element of the CAD-model it has to be pasted.

There are several other interesting facts for reconstructing the orientation of a camera:

- If the intrinsic parameters are not known the 3 coordinates of the projection center in the image coordinate system can be determined, together with the rotation matrix using a spatial resection.
- If the principle point is known the principle distance and the rotation matrix can be determined from two vanishing points.
- If the intrinsic parameters of the camera are known the rotation matrix can be determined from 2 straight image line segment pointing to one vanishing point and one straight line segment pointing to one of the other two vanishing points.

The last statements need a short elaboration, which also demonstrates the advantage of homogeneous coordinates and projective geometry.

Let the three endpoints $P_i(\boldsymbol{p}_i)$ and $Q_i(\boldsymbol{q}_i)$ of the line segments $l_i(\boldsymbol{n}_i)$ be represented as the normalized direction vectors

$$\boldsymbol{d}^T = \mathsf{N}(x\,y\,c) = \frac{(x\,y\,c)}{\sqrt{x^2+y^2+c^2}} \tag{5}$$

from the projection center $O$ to the points. It is assumed that the first two lines $l_1$ and $l_2$ point towards the first vanishing point $V_1$ and the third line $l_3$ points towards the second vanishing point $V_2$. Then the normal vectors $\boldsymbol{n}_i$ of the planes $\varepsilon_i = (l_i, O)$ through the lines and the projection centers is given by

$$\boldsymbol{n}_i = \mathsf{N}(\boldsymbol{p}_i \times \boldsymbol{q}_i), \qquad i = 1,2,3 \tag{6}$$

The 3 vanishing points $V_j(\boldsymbol{v}_j)$ and the rotation matrix are then easily determined by

$$\boldsymbol{v}_1 = \mathsf{N}(\boldsymbol{n}_1 \times \boldsymbol{n}_2) \qquad \boldsymbol{v}_2 = \mathsf{N}(\boldsymbol{v}_1 \times \boldsymbol{n}_3) \qquad \boldsymbol{v}_3 = \mathsf{N}(\boldsymbol{v}_1 \times \boldsymbol{v}_2) \qquad \boldsymbol{R} = (\boldsymbol{v}_1\;\boldsymbol{v}_2\;\boldsymbol{v}_3) \tag{7}$$

Eq. (7a) results from the intersection of the two lines. Eq. (7b) results from the fact that the vanishing points $V_2$ and $V_3$ and the projection center $O$ span a plane perpendicular to $\boldsymbol{v}_1$, thus $\boldsymbol{v}_1$ also represents the line through $V_2$ and $V_3$. Finally, eq. (7c) results from the orthogonality of the three directions. Using the three lines highlighted in fig. 4b yielded a rotation matrix which deviated from the one of the above mentioned clustering procedure by less than a degree. Only 6 crossproducts are required. A semiautomatic procedure for determining an approximation for the rotation matrix would provide the line segments overplayed on the image. Three mouse clicks on the image would be sufficient to obtain the rotations matrix.

### 3.4. Automatic Evaluation: Traffic Light Programs

Automatic vision tools require self diagnosis as vision modules always are a subpart in a larger system. The submodule is responsible for solving the task and reporting on its success or failure. R. Kroon coined the term *traffic light program* for a piece of software which is able to explicitly declare its result to be correct (green), incorrect (red) or partially correct (yellow), possibly indicating error sources.

Tools for achieving this goal have been developed in geodetic research over the past 30 years and cover a rich body of tools for evaluating the results of estimation procedures.[39,40] The Computer Vision community became aware of this problem and since a few years intensively discussed the issue of performance analysis.[41–43] It is not clear from the beginning that the classical statistical techniques for evaluating the testability of observations and the sensitivity of the results with respect to detectable errors can be transferred to the area of image analysis. First results have been reported by Sester et al.,[44] which later lead to the automatic model based orientation procedure[9] using man-made structures as control points,[45] and which demonstrated the possibility to use classical diagnostic tools for evaluating image analysis procedures.

The experiences can be directly transferred to all types of videometric procedures which then would significantly increase their reliability in use for non-specialists.

## 4. RECOGNITION FOR 3D-MODELING

Measuring instruments yield data, not information. Information is obtained by appropriate linking of the data with a model how to interprete the data. In interactive systems this model is not made explicit as the human operator is performing the image interpretation and, based on the corresponding descisions, measures the geometric attibutes of the object either indirectly, in the image, or directly, in 3D.

When automating the derivation of a symbolic description of the 3D-shape of objects from image or range data a generic scene model is required. We want to discuss the impact of this model on the 3D-reconstruction process and the accuracy of the result.

### 4.1. Segmentation: The role of geometric scene knowledge

The scene model needs to be detailed enough that an automatic interpretation procedure is able to infer the structure of the object from the iconic data. For highly structured scenes local constraints easily can be formalized and – in a first step – can be reduced to geoemtric features of the scene.

The surface of man made structures e. g. often can be described by a composition faces. This information is necessary to trigger automatic segmentation processes which can be viewed as a raster-vector transformation, also in the case of triangulated surfaces. Overviews and experimental comparisons for segmentation techniques for range images is given by Arman et al.,[46] Hoover et al.[47] and Leonardis.[48] A successful implementation has been described by Sequeira.[5]

The local constraints, which may require surface patches to be planes or quadrics, especially cylinders, cones or spheres, are needed to evaluate the local roughness of the measured data in order to detects and percisely locate patch boundaries. In case of a consistent use of these constraints the boundaries automatically should get the right geometry, e. g. in the case of two neighbouring planes the common boundary will then be straight. Besides this geometric knowledge also the topological contraints need to be considered, which is not standard in segmentation procedures, especially region growing algorithms are not able to explicitely model long boundaries.

it is an open question how these methods need to be modified or even rebuilt in order to meet their requirements.

### 4.2. 3D-Cite-Modeling: The role of semantic scene knowledge

It is not clear from the beginning that thematic or semantic knowledge could help image metrology. We therefore want to give an example that the automatic interpretation of the scene using semantic knowledge actually increases the accuracy of the reconstructed scene.[49]

Fig. 5 shows four image patches taken from digitized aerial images with known orientation parameters. The image scale is 1 : 5000, the pixel size 30 m, corresponding to a resolution of appr. 15 cm on the ground. The task is to reconstruct the 3D-shape of the building and classify its parts.

The procedure starts with automatically extracting image features (cf. fig. 6) using the same routines as above.[35,36] The system then identifies local image aggregates which may correspond to building edges, by analysing the neighbourhood of the extracted points. The hypotheses for image corners are then used in a matching process to reconstruct the 3D-corner, i. e. the 3D-position of the point and the orientations of the neighbouring edges. The result of this 3D-reconstruction is shown in the table. The empirical standard deviations of the three coordinates result from a least squares adjustment using the 4 observed points (8 observatios) and the 19 edges (38 observations) in the different images, some too short to be visible in fig. 7.

Now the scene knowledge is used to first classify the orientation of the edges to be either vertical, horizontal or oblique and to find possible symmetries. This classification leads to the interpretation of the 3D-corner, namely to be a symmetric gable point. The 4 constraints, the gable line to be horizontal and the other two lines to be symmetric in a vertical plane reduce the number of unknown to 5. The result of the adjustment is shown in the second line of the table. The value $\sigma_0$ states the internally derived precision of the observatios to be overestimated by appr. a factor 5. Its increas is statistically isignificant, proving the found constraints not to be wrong. The classificatio of the corner as a gable corner is the most probable interpretation, when using a Bayesian classificator. The decrease in unknowns leads to a reduction of the standard deviations of the 3D-coordinats up to a factor 2. This increase is the result of the use of the interpretation of the 3D-corner. Without the semantic knowledge the certainty of the classification of the edge orientations would be much less.

**Figure 5.** Four image patches of a man made structure

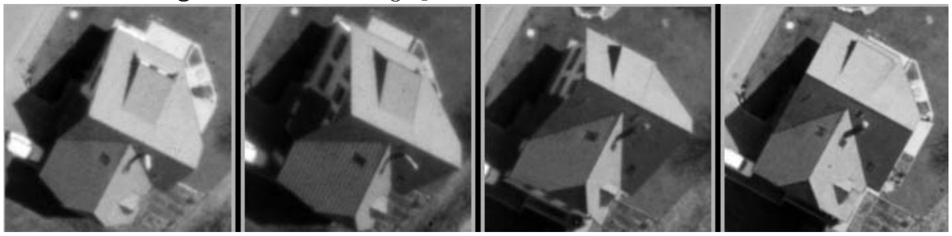
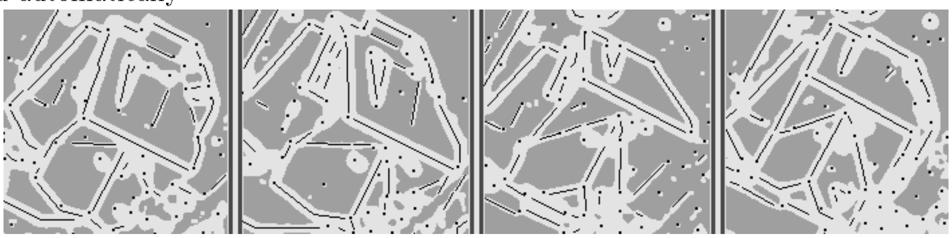


**Figure 6.** Automatically extracted features: points line segments and regions. The relations betwen the features are also extracted automatically



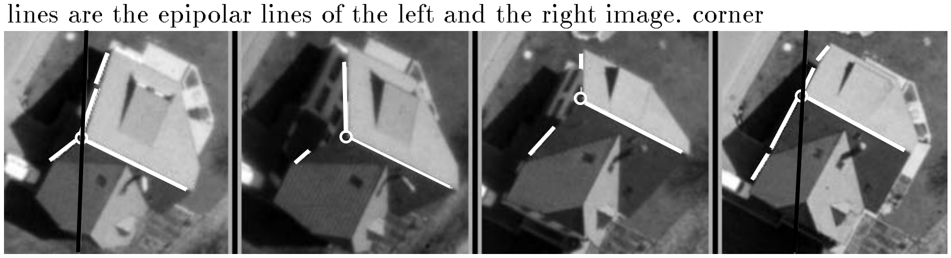
| Case | $n$ | $u$ | $\sigma_0$ [1] | $\sigma_x$ [$m$] | $\sigma_x$ [$m$] | $\sigma_z$ [$m$] |
|---|---|---|---|---|---|---|
| no constraints | 46 | 9 | 5.46 | 0.034 | 0.024 | 0.060 |
| with constraints | 46 | 5 | 5.52 | 0.019 | 0.016 | 0.032 |

When reconstruction scenes for virtual reality applications the increase in accuracy is less important as the consistency of reconstructed scene with the – usually not explicitely stated – model of the scene, door ways, for example, should show rectangular structure, even if the shortness of the 3D-edges and the accuracy of the sensor do not allow a reliable mensuration of the edge.

## 5. CONCLUSIONS

The impact of scene and assessment models for image metrology has been discussed. The focus on achieving highest possible accuracy has triggered research and development of highly sophisticated sensors. However, it needs to be embedded into a broader context of scene analysis where measurements can be just as inaccurate as tolerable for the specific task. This could enlarge the field of application of videometry significantly.

**Figure 7.** The original images overlayed with the features (white) used for matching and reconstruction of the gable. The black lines are the epipolar lines of the left and the right image. corner



## REFERENCES


1. B. G. Batchelor and P. F. Whelan, "Machine vision systems: Proverbs, principles, prejudices and priorities," in *Machine Vision Applications, Architectures, and Systems Integration III*, pp. 374–383, SPIE 2347, 1994. http://www.eeng.dcu.ie/ whelanp/proverbs/proverbs.html.
2. S. Abraham and T. Hau, "Towards autonomous high-precision calibration of digital cameras," in *Videometrics V*, SPIE 3174, July 1997.
3. P. Beardsley, P. Torr, and A. Zisserman, *3D Model Acquisition from Extended Image Sequences*, vol. 1065 of *Lecture Notes in Computer Science*, pp. 683–695. Springer-Verlag, 1996.
4. C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorisation method," *International Journal of Computer Vision* , pp. 137–154, 1992.
5. V. Sequeira and M. I. R. J. G. M. Gonçalves, "3d scene modelling from multiple range views," in *Videometrics IV*, S. El-Hakim, ed., pp. 114–127, SPIE 2598, 1995.
6. S. El-Hakim and J.-A. Beraldin, "Configuration design for sensor integration," in *Videometrics IV*, S. El-Hakim, ed., pp. 274–285, SPIE 2598, 1995.
7. H. Schewe, *Automatic Photogrammetric Measurement of Car Bodies*, pp. 47–55. Schriftenreihe Heft 12, Proceedings of the 41st Photogrammetric Week, Institut für Photogrammetrie, Universität Stuttgart, 1987.
8. F. Ackermann and P. Krzystek, "Match-t: Automatic mensuration of digital elevation models," in *Technical Seminar of the Sociedad Espanola de Cartografia Fotogrametria y Teledeteccion*, 43rd Photogrammetric Week, pp. 67–73, April 12 1991.
9. W. Schickler, "Feature matching for outer orientation of single images using 3-d wireframe controlpoints," in *Int. Arch. f. Photogrammetry and Remote Sensing, Vol. 29, III*, 1992.
10. H. Haggrén, *On System Development of Photogrammetric Stations for On-Line Manufacturing Control.* PhD thesis, Helsinki University of Technology, Acta Polytechnica Scandinavica, Civil Engineering and Building Construction Series No.97, 1992.
11. W. Bösemann, "The optical tube measurement system OLM photogrammetic methods used for industrial automation and process control," in *Intl. Archives of Photogrammetry and Remote Sensing*, pp. 55–58, IAP, Vol. XXXI,B5, 1996.
12. W. Förstner and L. Plümer, *Semantic Modeling and Extraction of Spatial Objects from Images and Maps*, Birkhäuser, 1997.
13. O. D. Faugeras, *What can be seen in three dimensions with an uncalibrated stereo rig?*, vol. 588 of *Lecture Notes in Computer Science*, pp. 563–578. Springer, 1992.
14. R. Hartley, *Estimation of Relative Camera Positions for Uncalibrated Cameras*, vol. 588 of *Lecture Notes in Computer Science*, pp. 579–587. Springer, 1992.
15. J. L. Mundy and A. P. Zisserman, *Geometric Invariance in Computer Vision*, MIT Press, 1992.
16. W. Förstner, "A future of photogrammetric research," *NGT Geodesia* **93**, pp. 372–383, August 1993.
17. K. Ikeuchi and T. Kanade, "Modeling sensor detectability and reliability in the configuration space for model-based vision," Tech. Rep. CMU-CS-87-144, Carnegie Mellon, July 1987.
18. W. Förstner and H. Werner, "Programmgesteuerte Deformationsanalyse," *ZFV* **111. Jahrgang**, pp. 236–247, June 1986.
19. H. Lehmkühler, *Die geodätische Deformationsanalyse als Mustererkennungsaufgabe.* PhD thesis, RWTH Aachen, D82, 1993.
20. P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley NY, 1987.
21. G. Vosselman and W. Förstner, "The precision of a digital camera," in *Intl. Archives of Photogrammetry and Remote Sensing*, IAP, Vol. 27,B, 1988.
22. M. R. Shortis, T. A. Clarke, and S. Robson, "Practical testing of the precision and accuracy of target image centring algorithms," in *Videometrics IV*, S. El-Hakim, ed., pp. 65–76, SPIE 2598, 1995.
23. J. W. Hill, "Dimensional measurement for quantized images," tech. rep., SRI Project 4391, 1983.
24. W. Förstner, "On the Geometric Precision of Digital Correlation," in *Int. Archives of Photogrammetry, Vol. 24-III, Helsinki*, J. Hakkarainen, E. Kilpelä, and A. Savolainen, eds., pp. 176–189, ISPRS Comm. III, jun 1982.

25. W. Förstner, "Uncertain spatial relationships and their use for object location in digital images," in *Robust Computer Vision - Tutorial Notes*, W. Förstner, R. M. Haralick, and B. Radig, eds., Institut für Photogrammetrie, Universität Bonn, 1992.
26. M. Berman and D. Culpin, "The statistical behaviour of some least square estimates of the centre and radius of a circle," *J. R. Statist. Soc. B* **48**(2), pp. 183–196, 1984.
27. N. Casott and J. Prenting, "Bestimmung des Auflösungsvermögens der CCD-Kamera JenaCam 43," tech. rep., Geodätisches Institut, Universität Bonn, 1997.
28. K. Voss and H. Suesse, "Invariant fitting of planar objects by primitives," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(1), pp. 80–84, 1997.
29. A. Shashua, *Trilinearity in visual recognition by alignment*, vol. 800 of *Lecture Notes in Computer Science*, pp. 479–484. Springer-Verlag, 1994.
30. R. Hartley, "A linear method for reconstruction from lines and points," in *Fifth International Conference on Computer Vision*, vol. 588 of *Lecture Notes in Computer Science*, pp. 882–887, IEEE, IEEE Computer Society Press, June 1995.
31. A. Krupnik, "Multiple-patch matching in the object space for aerotriangulation," Tech. Rep. 428, Department of Geodetic Science and Surveying, The Ohio State University, August 1994.
32. P. Krzystek, T. Heuchel, U. Hirt, and F. Petran, *A New Concept for Automatic Digital Aerial Triangulation*, pp. 215–224. Photogrammetric Week'95, Wichmann Verlag, 1995.
33. W. Mayr, *Aspects of Autoamtic Aerotriangulation*, pp. 225–234. Photogrammetric Week'95, Wichmann Verlag, 1995.
34. J. H. Lambert, "Anlage zur Perspektive," in *Schriften zur Perspektive*, J. H. Lambert, ed., Steck, Lüttke Verlag, Berlin, 1759.
35. W. Förstner, *A Framework for Low Level Feature Extraction*, pp. 383–394. LNCS 802, Springer, 1994.
36. C. Fuchs and W. Förstner, "Polymorphic grouping for image segmentation," in *Proc. 5th ICCV*, pp. 175–182, 1995.
37. M. Wydera, "Automatische Bestimmung von Fluchtpunkten," Master's thesis, Institut für Photogrammetrie, Universität Bonn, 1995.
38. M. Bank, "Entwicklung eines Merkmalszuordnungsverfahrens zur Orientierung von Bildern künstlicher Objekte," Master's thesis, Institut für Photogrammetrie, Universität Bonn, 1996.
39. W. Förstner, "Reliability Analysis of Parameter Estimation in Linear Models with Applications to Mensuration Problems in Computer Vision," *Computer Vision, Graphics & Image Processing* **40**, pp. 273–310, 1987.
40. W. Förstner, "Diagnostics and Performance Evaluation in Computer Vision," in *Proc. Performance versus Methodology in Computer Vision, NSF/ARPA Workshop, Seattle*, pp. 11–25, IEEE Computer Society, 1994.
41. R. M. Haralick, "DIALOGUE - Performance Characterization in Computer Vision," *CVGIP: Image Understanding* **60**(2), pp. 245–249, 1994.
42. W. Förstner, "The role of robustness in computer vision," in *Vision Milestones 95*, A. Pinz and W. Burger, eds., ch. 6, Österr. Gesellsch.. f. Künstliche Intelligenz, 1995.
43. W. Förstner, "10 pros and cons against performance characterization of vision algorithms," in *Workshop "Performance Characterisics of Vision Algorithms"*, M. C. B. Christensen H. I., Förstner W., ed., 1996.
44. M. Sester and W. Förstner, "Object Location Based on Uncertain Models," in *Proc. DAGM Symposium, Hamburg*, 1989.
45. T. Läbe and K.-H. Ellenbeck, "3D-wireframe models as ground control points for the automatic exterior orientation," in *Proceedings ISPRS Congress, Comm. II, Vienna, IAP Vol. XXXI*, 1996.
46. F. Arman and J. K. Aggarwal, "Model-based object recognition in dense-range images – a review," *ACM Computing Surveys* **25**(1), pp. 5–43, 1993.
47. A. Hoover, G. Jean-Baptiste, X. Jiang, P.J.Flynn, H. B. nad D. Goldgof, K. Bowyer, D. Eggert, A. Fitzgibbon, and R. Fisher, "An experimental comparison of range imaghe segmentation algorithms," *IEEE T-PAMI* **18**(7), pp. 673–689, 1996.
48. A. Leonardis, *Image Analysis Using Parametric Models.* PhD thesis, Faculties of Electrical Engineering and Computer Science, University of Ljubljana, 1993.
49. F. Lang and W. Förstner, "Surface reconstruction of man-made structures using polymorphic mid-level features and generic scene knowledge," in *ISPRS Symposium, Vienna*, 1996.