Sequential Estimation in Robot Vision*

Armin Gruen and Thomas P. Kersten

Abstract

Highly time-constrained robot vision applications require a careful tuning and optimized interaction of a system's hardware components, algorithmic complexity, software engineering, and task performance. The high accuracy processing of full-frame image sequences for image analysis and object space feature positioning is very time consuming. In both of these processes, sequential estimation algorithms offer valuable alternatives to simultaneous approaches. This paper introduces an efficient estimation algorithm based on Givens transformations for use in point positioning and updating camera orientation data. In a test, an easy-to-use standard video camera has been applied for image frame generation. The results of camera calibration and an accuracy test using a 3D testfield are presented. The computing times of sequential point positioning and camera orientation are given and in part compared to the values for the simultaneous adjustment. This clearly indicates the superior performance of the sequential procedure.

Introduction

Image sequences play an important role in photogrammetry, machine vision, and robot vision. While in classical photogrammetry, especially in aerial applications, data acquisition and processing is largely separated, this is no longer the case in modern applications where non-photographic sensor technology and digital processing techniques are employed. Fast methods for data reduction are required, in particular, in highly time-constrained robotics applications, but are also very often of advantage in less time-critical machine vision and digital photogrammetric projects. The classical data reduction process consists of two major stages: image measurement and three-dimensional (3D) point positioning. These process components are in general separated from each other. In each case, simultaneous algorithms can be reformulated into sequential form for better time performance.

In image processing, well-known sequential formulations exist for incremental convolution operations (used in linear filtering, resampling, image pyramid generation, etc.); in image analysis, they are applied in the pixel location transformations in orthophoto production (Baltsavias *et al.*, 1991) and in form of the Kalman filter in the tracking of line segments in image space (Deriche and Faugeras, 1990).

A well known example is that of on-line triangulation using sequential estimation techniques in point positioning with aerial photographs. Here the computational procedure of on-line bundle triangulation is closely tied to the image coordinate measurement process involving a human operator. The main purpose of this fast sequential estimation is that of blunder detection at an early stage of the measurement process with the utilization of quick remeasurement possibilities and better blunder control capabilities. Important characteristics of this application are, on the one hand, the constantly varying size of the state vector ("solution vector" in least-squares adjustment terminology) of bundle adjustment consisting of the exterior orientation parameters of photographs, the object point parameters, and, possibly, additional parameters for self-calibration. On the other hand, the full covariance matrix of all system parameters is, if at all, only needed at the termination of the process. A third distinctive characteristic is the high and typically sparse patterns of the matrices involved in the estimation procedure (design matrix of observation equations and normal equation matrices of least squares). Given these system characteristics, a number of sequential estimation algorithms have been compared to each other in the past. First, the Triangular Factor Update (TFU) algorithm, which updates directly the upper triangle of the reduced normal equations, was found to perform much better than the Kalman form of updating, both in terms of computing times and storage requirement (Gruen, 1982; Wyatt, 1982). Later, the Givens transformations were found to be superior, in general, to the TFU (Runge, 1987; Holm, 1989), both in computational performance and in the ease of mechanization and software implementation. In the meantime, the Givens algorithm has been implemented in a number of systems (Edmundson, 1991; Kersten *et al.*, 1992). Already in the mid 80s, Gruen (1985a) envisioned semi-automatic or fully automatic digital real-time triangulation systems for the future. We argue nowadays that machine vision and, in particular, robot vision could draw substantial advantages from these sequential approaches.

This fact has been obviously acknowledged by the computer vision community, where, among others, two recent developments are of particular interest. In Matthies *et al.* (1989), the Kalman filter is used to estimate a depth map from image sequences. Typical for this approach is that depth estimates and uncertainty values are computed at each pixel of minified (256 by 256 pels) video frames and that these estimates are refined incrementally over time. The good performance of the Kalman filter is due to the fact that only one parameter (depth value) constitutes the state vector and is updated. Errors in orientation and calibration of the video frames and thus correlations between different elements of the depth map are not considered. Also, only lateral motion of the CCD camera is assumed.

*Presented at the ISPRS Congress, Commission V, Washington, D.C., 2-14 August 1992.

Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology, ETH-Hoenggerberg, CH-8093, Zurich, Switzerland.

Photogrammetric Engineering & Remote Sensing,
Vol. 61, No. 1, January 1995, pp. 75–82.

0099-1112/95/6101-75$3.00/0

Zhang and Faugeras (1990) use the Kalman update mechanism to track object motion in a sequence of stereo frames. They track object features, called "tokens" (line segments for example), in 3D space from frame to frame and estimate the motion parameters of these tokens in a unified way (they also integrate a model of motion kinematics). Their state vector consists, thus, of three angular velocity, three translational velocity, and three translational acceleration parameters of each object, and, in addition, six line segment parameters for each 3D line representation. As before, the tokens are treated independently here, which allows the state vector for estimation to remain small in size and may result in straightforward parallelization for any number of tokens.

Although we are aware that very fast (e.g., a video rate of 25 Hz) solutions to tracking problems with affordable computer hardware still require substantial simplification of the measurement problem at hand, we nevertheless present in this paper the general solution for sequential point positioning, based on the bundle solution. The solution is object-point based. Other features may be derived from these 3D point measurements. Any simplification in measurement arrangement, as for instance non-moving sensors, may readily be derived from the general concept. Our solution may include self-calibration parameters for systematic error modeling and statistical tests for blunder detection. The sequential estimation procedure applies Givens transformations for updating of the upper triangle of the reduced normal equations (Blais, 1983; Gruen, 1985a). We use Givens transformations as opposed to a Kalman update because in robotics the covariance update of the parameter vector is not required at every stage and then, if at all, only at relatively sparse increments. Moreover, the varying size of the parameter vector (addition and deletion of new object points, addition of frame exterior orientation parameters) leads to very poor computational performance of the Kalman filter.

The approach presented here treats only the 3D point positioning problem in a sequential mode. A combination of sequential algorithms in 2D image measurement and 3D object point positioning within one unique system is feasible and meaningful if, for instance Multiphoto Geometrically Constrained (MPGC) image matching, which delivers object point coordinates simultaneously, is executed for full frames in a pixel-by-pixel (iconic) mode. Also, a complete bundle solution with integrated image matching could be based on this concept. Another generalization is possible if moving objects are included in the system.

In the second section, a brief description of the Givens transformations, as applied to sequential bundle triangulation with static object points, is given. The third section presents a test example using real image data produced with an arbitrarily moving video camera over a 3D testfield.

Sequential Estimation in the Bundle Adjustment with Givens Transformations

In this section we will present in a concise fashion the formulae of Givens transformations as applied to sequential estimation in bundle systems. For a more comprehensive treatment, compare Blais (1983), Gruen (1985a), Runge (1987), and Holm (1989). In the following, our functional model for the bundle adjustment will be set up without the inclusion of parameters for self-calibration. An extension by these parameters is straightforward and does not alter the considerations and conclusions presented here.

Least-Squares Approach for Estimation

The Gauss-Markov model is the estimation model most widely used in photogrammetric linear or linearized estimation problems. An observation vector $\mathbf{l}$ of dimension $n \times 1$ is functionally related to a $u \times 1$ parameter vector $\mathbf{x}$ through

$$\mathbf{l} - \mathbf{e} = \mathbf{A}\mathbf{x} \tag{1}$$

The design matrix $\mathbf{A}$ is an $n \times u$ matrix with $n \geq u$ and Rank $(A) = u$. There is no need to work with rank-deficient design matrices in on-line triangulation. Rank deficient systems, caused by missing observations, generally do not allow for a comprehensive model check. Observations should be accumulated until the system is regular and can be solved using standard techniques. For rank deficiency caused by incomplete datum, see Gruen (1985a). Sequential least-squares estimation with pseudo-inverses is very costly (compare Boullion and Odell (1971), p. 50 ff). The vector $\mathbf{e}$ represents the true errors. With the expectation $E(\mathbf{e}) = 0$ and the dispersion operator D, we get

$$E(\mathbf{l}) = \mathbf{A}\mathbf{x}, \tag{2a}$$

$$D(\mathbf{l}) = C_{ll} = \sigma_0^2\, \mathbf{P}^{-1}, \text{ and} \tag{2b}$$

$$D(\mathbf{e}) = C_{ee} = C_{ll}. \tag{2c}$$

The estimation of $\mathbf{x}$ and σ_0^2 is usually attempted as unbiased, minimum variance estimation performed by means of least squares, and results in

$$\text{parameter vector } \hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{P}\mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}\mathbf{l}, \tag{3a}$$

$$\text{residual vector } \mathbf{v} = \mathbf{A}\,\hat{\mathbf{x}} - \mathbf{l}, \text{ and} \tag{3b}$$

$$\text{variance factor } \hat{\sigma}_0^2 = \frac{\mathbf{v}^T\mathbf{P}\mathbf{v}}{r},\ r = n - u. \tag{3c}$$

The architecture of $\mathbf{A}$ is determined by the type of triangulation method used. As explained previously, we chose the bundle method for the purpose of generality and rigidity.

For bundle adjustment, Equation 1 can be written as

$$-\mathbf{e} = \mathbf{A}_1\,\mathbf{x} + \mathbf{A}_2\,\mathbf{t} - \mathbf{l};\ \mathbf{P} \tag{4a}$$

where

$\mathbf{x}$ is the vector of object point coordinates;
$\mathbf{t}$ is the vector of orientation elements;
$\mathbf{A}_1$ and $\mathbf{A}_2$ are the associated design matrices; and
$\mathbf{e}$, $\mathbf{l}$, and $\mathbf{P}$ are the true error vector, constant vector, and weight matrix for image point observations, respectively.

$\mathbf{x}$ and $\mathbf{t}$ are considered here as unconstrained (free) parameters. If observations are available for some or all of the object point coordinates, a second system of observation equations is added, that is,

$$-\mathbf{e}_c = \mathbf{I}\mathbf{x} - \mathbf{l}_c;\ \mathbf{P}_c. \tag{4b}$$

Similarly, observations for the orientation elements would add

$$-\mathbf{e}_t = \mathbf{I}\mathbf{x} - \mathbf{l}_t;\ \mathbf{P}_t. \tag{4c}$$

The least-squares principle, applied to Equations 4a, 4b, and 4c leads to the combined minimum

$$\mathbf{v}^T\mathbf{P}\mathbf{v} + \mathbf{v}_c^T\,\mathbf{P}_c\,\mathbf{v}_c + \mathbf{v}_t^T\mathbf{P}_t\,\mathbf{v}_t \Rightarrow Min. \tag{4d}$$

For the purpose of simplicity and without loss of generality, we will operate in the following derivations only with the reduced minimum principle

$$\mathbf{v}^T\mathbf{P}\mathbf{v} \Rightarrow Min,$$

that is, we will consider only Equation 4a as observation equations.

The resulting normal equations are of the form

$$\mathbf{N}\begin{bmatrix}\hat{\mathbf{x}}\\ \hat{\mathbf{t}}\end{bmatrix} = \begin{bmatrix}\mathbf{N}_{xx} & \mathbf{N}_{xt}\\ \mathbf{N}_{xt}^T & \mathbf{N}_{tt}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{x}}\\ \hat{\mathbf{t}}\end{bmatrix} = \begin{bmatrix}\mathbf{l}_x\\ \mathbf{t}_t\end{bmatrix}, \tag{5}$$

with

$$\begin{aligned}\mathbf{N}_{xx} &= \mathbf{A}_1^T\mathbf{P}\mathbf{A}_1, \quad \mathbf{l}_x = \mathbf{A}_1^T\mathbf{P}\mathbf{l}\\ \mathbf{N}_{xt} &= \mathbf{A}_1^T\mathbf{P}\mathbf{A}_2, \quad \mathbf{l}_t = \mathbf{A}_2^T\mathbf{P}\mathbf{l}\\ \mathbf{N}_{tt} &= \mathbf{A}_2^T\mathbf{P}\mathbf{A}_2.\end{aligned} \tag{6}$$

N is further assumed to be regular. In an off-line environment, Equation 6 is usually solved by applying Gauss or Cholesky factorization. The former can formally be described as an LU factorization, decomposing **N** into a product of lower and upper triangular matrices **L** and U, i.e.,

$$\mathbf{L}\mathbf{U}\begin{bmatrix}\hat{\mathbf{x}}\\ \hat{\mathbf{t}}\end{bmatrix} = \begin{bmatrix}\mathbf{l}_x\\ \mathbf{l}_t\end{bmatrix} \tag{7}$$

or, with $L = \mathbf{U}^T\mathbf{D}$ (**D** is a diagonal matrix), in the alternate formulation

$$\mathbf{U}^T\mathbf{D}\mathbf{U}\begin{bmatrix}\hat{\mathbf{x}}\\ \hat{\mathbf{t}}\end{bmatrix} = \begin{bmatrix}\mathbf{l}_x\\ \mathbf{l}_t\end{bmatrix}. \tag{8}$$

After the reduction of the right hand side, the solution vector is computed from

$$\mathbf{U}\begin{bmatrix}\hat{\mathbf{x}}\\ \hat{\mathbf{t}}\end{bmatrix} = \mathbf{L}^{-1}\begin{bmatrix}\mathbf{l}_x\\ \mathbf{l}_t\end{bmatrix} \tag{9}$$

by back-substitution.

In photogrammetric triangulation, the factorization is usually done as a stepwise procedure, stopping the reduction of **N** right before it enters what was originally the $\mathbf{N}_{tt}$ matrix. This procedure leads to the pre-reduced normals $\mathbf{N}_R$, i.e.,

$$\mathbf{N}_R\hat{\mathbf{t}} = \mathbf{l}_R, \tag{10}$$

with $\mathbf{N}_R = \mathbf{N}_{tt} - \mathbf{N}_{xt}^T\mathbf{N}_{xx}^{-1}\mathbf{N}_{xt}$ and

$$\mathbf{l}_R = \mathbf{l}_t - \mathbf{N}_{xt}^T\mathbf{N}_{xx}^{-1}\mathbf{l}_x.$$

$\mathbf{N}_R$ is finally factorized to an upper triangle $\mathbf{N}_{RR}$ and $\hat{\mathbf{t}}$ is obtained by back-substitution from

$$\mathbf{N}_{RR}\hat{\mathbf{t}} = \mathbf{l}_{RR}. \tag{11}$$

The mechanization of this off-line factorization algorithm takes advantage of the fact that $\mathbf{N}_{xx}$ is a block-diagonal matrix with 3 by 3 submatrices along the diagonal. Therefore, the reduction of the point coordinates can be done on a "point by point" basis, leaving the structure of the $\mathbf{N}_{xx}$ and $\mathbf{N}_{xt}$ matrices unchanged, i.e., producing no new fill-ins in those matrices. This particular feature, based on the structure of $\mathbf{N}_{xx}$, is the key to a successful application of the Triangular Factor Update technique in on-line triangulation.

Assuming a sequential process and interpreting Equation 4a as the status of the measurement system at stage $k-1$ of the process, we get the following system if one or more image coordinate observations are added, including new parameters $\mathbf{x}_{(k)}$ and $\mathbf{t}_{(k)}$:

$$-\mathbf{e} = \mathbf{A}_1\mathbf{x} + \mathbf{A}_2\,\mathbf{t} - \mathbf{l};\ \mathbf{P}$$

$$-\mathbf{e}_{(k)} = \mathbf{A}_{1(k)}\begin{bmatrix}\mathbf{x}\\ \mathbf{x}_{(k)}\end{bmatrix} + \mathbf{A}_{2(k)}\begin{bmatrix}\mathbf{t}\\ \mathbf{t}_{(k)}\end{bmatrix} - \mathbf{l}_{(k)};\ \mathbf{P}_{(k)} \tag{12}$$

The updated normal equations of the stage k are of the form

$$\dot{\mathbf{N}}\begin{bmatrix}\hat{\dot{\mathbf{x}}}\\ \hat{\dot{\mathbf{t}}}\end{bmatrix} = \begin{bmatrix}\dot{\mathbf{l}}_x\\ \dot{\mathbf{l}}_t\end{bmatrix}, \tag{13}$$

with

$$\hat{\dot{\mathbf{x}}} \triangleq \begin{bmatrix}\hat{\mathbf{x}}\\ \hat{\mathbf{x}}_{(k)}\end{bmatrix}, \quad \hat{\dot{\mathbf{t}}} \triangleq \begin{bmatrix}\hat{\mathbf{t}}\\ \hat{\mathbf{t}}_{(k)}\end{bmatrix}$$

and

$$\dot{\mathbf{N}} = \begin{bmatrix}\dot{\mathbf{N}}_{xx} & \dot{\mathbf{N}}_{xt}\\ \dot{\mathbf{N}}_{xt}^T & \dot{\mathbf{N}}_{tt}\end{bmatrix}, \quad \begin{aligned}\dot{\mathbf{l}}_x &= \mathbf{l}_x^{(0)} + \mathbf{A}_{1(k)}^T P_{(k)}\mathbf{l}_{(k)},\\ \dot{\mathbf{l}}_t &= \mathbf{l}_t^{(0)} + \mathbf{A}_{2\,(k)}^T\,\mathbf{P}_{(k)}\,\mathbf{l}_{(k)}\end{aligned}$$

$$\begin{aligned}\dot{\mathbf{N}}_{xx} &= \mathbf{N}_{xx}^{(0)} + \mathbf{A}_{1\,(k)}^T\mathbf{P}_{(k)}\mathbf{A}_{1(k)}^T,\\ \dot{\mathbf{N}}_{xt} &= \mathbf{N}_{xt}^{(0)} + \mathbf{A}_{1(k)}^T\mathbf{P}_{(k)}\mathbf{A}_{2(k)}^T, \text{ and}\\ \dot{\mathbf{N}}_{tt} &= \mathbf{N}_{tt}^{(0)} + \mathbf{A}_{2(k)}^T\mathbf{P}_{(k)}\mathbf{A}_{2(k)}^T.\end{aligned}$$

The superscripts (0) indicate that, if new parameters $\mathbf{x}_{(k)}$ and $\mathbf{t}_{(k)}$ are added, the column/row spaces of the original $\mathbf{N}_{xx}$, $\mathbf{N}_{xt}$, and $\mathbf{N}_{tt}$ matrices have to be extended by zero vectors and the row spaces of the original vectors and by zero elements accordingly.

The updating of the $k-1$ stage normals can be described as

$$(\mathbf{N}+\Delta\mathbf{N})\begin{bmatrix}\hat{\dot{\mathbf{x}}}\\ \hat{\dot{\mathbf{t}}}\end{bmatrix} = \begin{bmatrix}\mathbf{l}_x + \Delta\mathbf{l}_x\\ \mathbf{l}_t + \Delta\mathbf{l}_t\end{bmatrix}. \tag{14}$$

The addition of the term $\Delta\mathbf{N}$ to the $k-1$ normals will result in alterations of the matrix factors **L** and **U**, i.e.,

$$(\mathbf{U}+\Delta\mathbf{U})\begin{bmatrix}\hat{\dot{\mathbf{x}}}\\ \hat{\dot{\mathbf{t}}}\end{bmatrix} = (\mathbf{L}+\Delta\mathbf{L})^{-1}\begin{bmatrix}\mathbf{l}_x+\Delta\mathbf{l}_x\\ \mathbf{l}_t+\Delta\mathbf{l}_t\end{bmatrix}. \tag{15}$$

Sequential Treatment with Givens Transformations

Sequential estimation with orthogonal transformations using QR decomposition is described in Lawson and Hanson (1974). Both additions/deletions of column and row vectors of the **A** matrix are discussed there. Householder transformations as well as Givens rotations are used.

Blais (1983) recommended the application of Givens rotations for the sequential treatment of surveying and photogrammetry networks. Our approach uses the estimation model (Equation 4a). Instead of obtaining the updated upper triangular matrix (Equation 15) by means of Gauss factorization of the normal equations, it applies Givens transformations directly to the upper triangular matrix **U** of the previous stage. At stage $k-1$ the reduced system (Equation 9) takes the form

$$\mathbf{U}\begin{bmatrix}\hat{\mathbf{x}}\\ \hat{\mathbf{t}}\end{bmatrix} = \mathbf{L}^{-1}\begin{bmatrix}\mathbf{l}_x\\ \mathbf{l}_t\end{bmatrix} = \mathbf{d}.$$

Adding one observation equation, including a set of new parameters **y**, to this system results in stage k and gives (with $\mathbf{P}_{(k)} = \mathbf{I}$)

$$\begin{bmatrix} U \mid 0 \\ \mathbf{a}^T_{(k)} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{t} \\ \hline \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \hline \mathbf{l}_{(k)} \end{bmatrix} \tag{16}$$

in which

$\mathbf{y}$ is the new parameter vector of length p,

$\mathbf{a}^T_{(k)}$ is the row vector with the coefficients of the new observation equation, and

$\mathbf{l}_{(k)}$ is the right hand side of the new observation equation.

Applying a series of orthogonal Givens transformations

$$\mathbf{G}=\mathbf{G}_n\mathbf{G}_{n-1} \ldots \mathbf{G}_1 \text{ (n is the total number of system parameters)} \tag{17}$$

to Equation 16 results in

$$\mathbf{G}\begin{bmatrix} \mathbf{U}\ 0 \\ \hline 0\ 0 \\ \hline a^T_{(k)} \end{bmatrix} \begin{matrix} \}n-p \\ \}p \\ \}1 \end{matrix} = \begin{bmatrix} \dot{\mathbf{U}} \\ \hline \mathbf{0} \end{bmatrix} \begin{matrix} \}n \\ \\ \}1 \end{matrix}, \tag{18a}$$

$$G\begin{bmatrix} d \\ \hline 0 \\ \hline \mathbf{l}_{(k)} \end{bmatrix} \begin{matrix} \}n-p \\ \}p \\ \}1 \end{matrix} = \begin{bmatrix} \dot{\mathbf{d}} \\ \\ \tilde{\mathbf{l}}_{(k)} \end{bmatrix} \begin{matrix} \}n \\ \\ \}1 \end{matrix}, \tag{18b}$$

The updated solution vector can be found by back-substitution into

$$\dot{\mathbf{U}} \begin{bmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{t}} \\ \hat{\mathbf{y}} \end{bmatrix} = \dot{\mathbf{d}}. \tag{19}$$

The sparsity patterns of both $\mathbf{U}$ and $\mathbf{a}^T_{(k)}$ can be exploited advantageously in order to speed-up computations.

If a covariance matrix of the parameters has to be updated, essentially the same approach can be followed as for the updated parameters. Another option is to derive it from the upper triangle $\mathbf{U}$ using Equation 14 in Gruen (1985a).

Methods for the deletion of observations and the addition and deletion of parameters are described in Golub (1969) and Lawson and Hanson (1974). Some of these methods fit nicely into the mechanization of the Givens approach. Deletion of observations can be handled by introducing these observation equations with negative weights into the standard format (Equation 16). Complex arithmetic is avoided in computations.

For the deletion of parameters, one simply cuts out the corresponding columns of the upper triangle $\mathbf{U}$ and transforms the remaining matrix to upper triangular form with Givens matrices. The transformation of vector $\mathbf{d}$ is also necessary.

The variance factor can be updated either through explicit computation in Equation 3c after the "new" residuals have been determined or through a sequential approach using the Givens transformations. Lawson and Hanson (1974, page 6) have shown that

$$\Omega = (\mathbf{v}^T\mathbf{P}\mathbf{v})^{1/2} = \mathbf{d}_2^T\mathbf{d}_2 \tag{20}$$

with d_2 being derived from the factorization of the observation Equation 4a as

$$\mathbf{G}\mathbf{P}^{1/2}[\mathbf{A}_1\ \mathbf{A}_2] = \begin{bmatrix} \mathrm{U} \\ 0 \end{bmatrix},\ \mathbf{G}\mathbf{P}^{1/2}\,\mathbf{l} = \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix}. \tag{21}$$

Hence, $\hat{\sigma}_0^2 = \Omega^2/r$ can easily be derived from the lower portion $\mathbf{d}_2$ of the transformed right hand side of the $(k-1)^{\text{th}}$ stage of the system. The sequential updating of Ω can be achieved by simply adding Ω to $\mathbf{d}$ and updating Ω with Givens transformations (Gentleman, 1973): i.e.,

$$\mathbf{G}\begin{bmatrix} \mathrm{U} \\ 0 \\ \mathbf{a}^T_{(k)} \end{bmatrix} = \begin{bmatrix} \dot{\mathrm{U}} \\ 0 \\ 0 \end{bmatrix},\ \mathbf{G}\begin{bmatrix} \mathbf{d} \\ \Omega \\ \mathbf{l}_{(k)} \end{bmatrix} = \begin{bmatrix} \dot{\mathbf{d}} \\ \dot{\Omega} \\ 0 \end{bmatrix}. \tag{22}$$

For blunder detection, which is particularly important in automated measurement systems, Baarda's data snooping has become standard. The computation of the related test criteria $\mathbf{w}_i = -\mathbf{v}_i/\sigma_{\mathbf{v}_i}$, $(i = 1, \ldots, n)$, requires the computation of both the residual vector $\mathbf{v}$ and the diagonal elements of the $\mathbf{Q}_{vv}$-matrix. In Gruen (1985a) it has been shown how the full $\mathbf{Q}_{vv}$-matrix can be efficiently computed or updated both with the Givens and the TFU approach. In practical operations, we clearly prefer the method of "unit observation vector" (Gruen, 1982), because in the process of on-line triangulation only a relatively small number of selected residuals have to be tested at any given stage. An operational procedure for blunder detection and deletion of gross erroneous observations in the case of non-diagonal-dominant $\mathbf{Q}_{vv}\mathbf{P}$-matrices, and if the suspicion exists that more than one blunder is in the data at a given time, was suggested in Gruen (1985b). After each deletion of an observation, the remaining residuals and the estimated variance factor are updated in a recursive fashion and tested again in order to get rid of the influence of the rejected observation.

Practical Example

The software package OLTRIS (On-Line Triangulation System; see Kersten *et al.* (1992) for a description), which was developed for the U.S. DoD (Department of Defense), is used for the computation of the sequential estimation operations in the practical example of this section.

Figures 1, 2a, and 2b show our 3D laboratory testfield which served as the test object to be imaged by our (simulated) robot, equipped with one CCD camera. In fact, instead of a robot, a human operator "filmed" an image sequence with a JVC video camera GR-S77E (S-VHS).

This "amateur" video camera (Figure 3) was intentionally used instead of an industrial CCD camera, because of its ease of operation (e.g., viewfinder for optimal object coverage and internal compact video cassette for immediate data storage of very long image sequences). Table 1 shows some of the technical specifications of this camera.

Image Frame Generation

The recording "flight" path of the video camera is illustrated in Figures 2a and 2b. The sequence of the 3D testfield was recorded in two strips, moving the "robot" parallel to the testfield at an approximate distance of 3.6 metres from the wall. While recording the images, the auto focus was switched off and the camera was focused at infinity. With the focal length of the camera fixed at 8.5 mm, the depth-of-field can be assumed sufficient for sharp imaging of the object. Fifty-three seconds of the sequence have been chosen for digitization. The imagery was digitized with a VideoPix

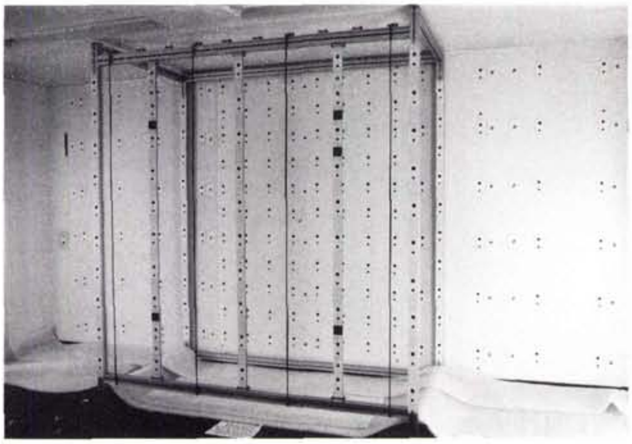


Figure 1. Three-dimensional test field used for on-line triangulation.

framegrabber on a SPARCstation 1+ (Sun Microsystems). The generated image frames were pre-processed with a low-pass filter (3 by 3 average). The effective size of each digitized image was 720 (H) by 575 (V) pixels. Altogether, 90 image frames were generated, giving a rate of 1.8 images per second of the sequence or one digitized image every 0.6 seconds. Due to blurring effects caused by image motion, two image frames were left out of the digitized sequence. Figure 4 shows three frames of the complete sequence (enhanced with a Wallis filter for illustration). The original visual quality of the frames is not very good. Radiometric and thus geometric distortions due to motion blur, analog video cassette storage, and frame grabbing with PLL line synchronization are visible if imaged at larger scales.

Camera Calibration

Before measuring image coordinates and processing data in OLTRIS, the video camera was calibrated. In the calibration, additional parameters including parameters of interior orientation, x-scale factor, shear, and radial and decentering distortion were determined. Investigations into the calibration of CCD cameras are described by Beyer (1992). The respective software has also been used here.

In addition to the test sequence, images were acquired for calibrating the JVC. The 3D testfield was imaged from four different camera positions. The pixel coordinates of the test-field targets were determined by least-squares template matching (LSTM), while reference coordinates for the targets were obtained by theodolite measurements. Measuring some well-distributed points in the four images yielded sufficiently precise approximations for the exterior orientation of the four images by resection in space. Using these data and the known object point coordinates, approximate image coordinates could be computed. These were used as initial values for automatic least squares template matching of 130 points in each image. In this test, LSTM was capable of measuring seven targets per second including screen display with an average precision of 0.33 µm (1/33 pixels) in x and 0.29 µm (1/35 pixels) in y.

The observations were processed in a bundle adjustment with self-calibration. The measurements and adjustment were performed in DEDIP (Development Environment for Digital Photogrammetry; Beyer, 1987), which is a part of the Digital Photogrammetric Station DIPS II (Gruen and Beyer, 1990). The results from the bundle adjustment and the comparison with check points, as an independent verification of the accuracy, are shown in Table 2. Version 1 summarizes the results of the calibration with a minimal control datum (three control points on the wall). In version 2, eight well- distributed control points on the wall and test-field frame were used in the adjustment. The empirical accuracy measures (μ_X, μ_Y, μ_Z), which are computed after a spatial similarity transformation onto their check-point coordinates, show that an accuracy of better than one millimetre was obtained. A more accurate calibration of this video camera with a larger number of image frames and a different camera configuration was performed by Beyer *et al.* (1992).

An accuracy on the order of 1/10th of the pixel spacing in image space could be achieved. The camera constant was determined as c = 8.62 mm, and the pixel spacing as 9.1 µm (H) by 8.3 µm (V). The curve of radial distortion of the JVC is illustrated in Figure 5. The 6.4- by 4.8-mm^2 sensor of the JVC is affected by a maximum distortion of −71 µm at the sensor corners.

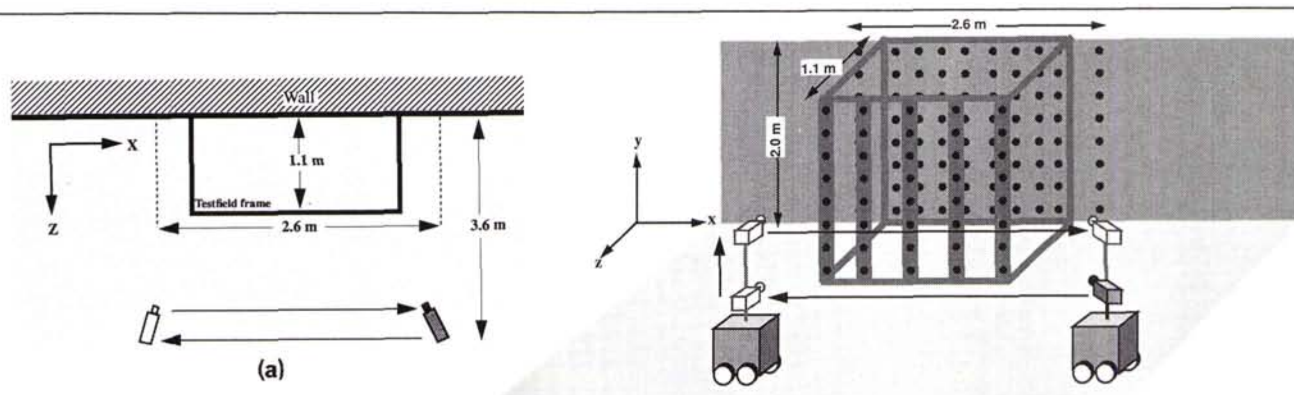


Figure 2. **(a)** Dimensions of the test field and camera path (planview). **(b)** Dimensions of the testfield and illustration of the camera path used for acquiring the test sequence.

Figure 3. JVC video camera GR-S77E.

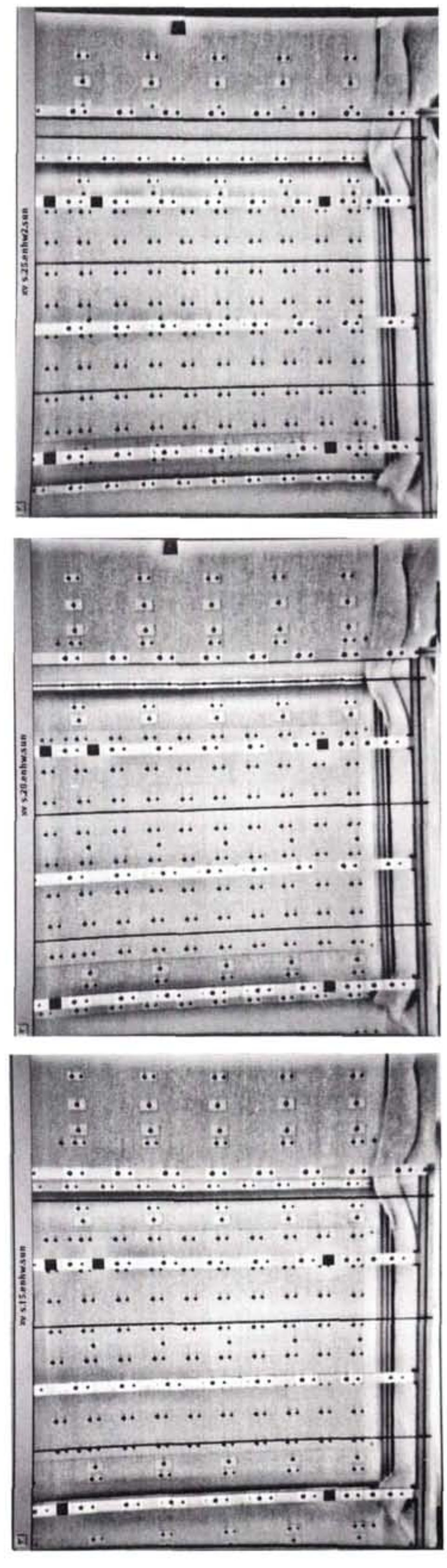


Figure 4. Three frames (numbers 10, 15, and 20) of the video sequence (enhanced with a Wallis filter for illustration).

On-Line Triangulation

In OLTRIS, the image sequence was triangulated to demonstrate the performance and capability of sequential adjustment for point positioning purposes. As mentioned earlier, the triangulation was processed without self-calibration. The image coordinates for the object points in the 88 images were determined in a similar fashion as described above for the camera calibration. Known data at the start of the triangulation included the station orientation data of the first image (introduced as initial values) and five distributed object points of the test field which defined the datum. After including a new frame into the triangulation process, at least three points have to be measured to compute the approximate values of the exterior orientation of the "current" camera position. These orientation values of each consecutive image in the sequence were computed by resection in space using the orientation data of the preceding image as initial values. In each image, between 79 and 146 points were measured. A total of 166 different object points in the test field were used. In total, 20,860 observations (object and 20,362 image point coordinates) were processed, with a maximum number of 1026 unknowns to be determined in the bundle adjustment. The path of the video camera for the test sequence is plotted in Figure 6. The lower line represents the estimated path (i.e., exterior orientation of the 88 images) as determined by OLTRIS (with sequential estimation and simultaneous adjustment inbetween). The upper line indicates the "path" as estimated in a (simultaneous) bundle adjustment with self calibration in DEDIP. The mean of the differences between the two paths is 4.5 mm in the x-direction,

TABLE 1. RELEVANT TECHNICAL SPECIFICATIONS OF THE JVC VIDEO CAMERA GR-S77E

JVC video camera GR-S77E
Super VHS System for record and play mode
High resolution 1/2-inch CCD Chip (420,000 pixels)
Focal length 8.5–68 mm, 8× zoom lens
Auto focus
Variable electronic shutter 1/50, 1/250, 1/500, 1/1000 sec
Weight 1.2 kg

TABLE 2. RESULTS OF THE BUNDLE ADJUSTMENT WITH SELF-CALIBRATION FOR THE VIDEO CAMERA CALIBRATION

							Precision from adjustment					Accuracy from check points[1]				
							Object space [mm]			Image space [µm]		Object space [mm]			Image space [µm]	
Ver	Im	AP	Co	Ch	r	$\hat{\sigma}$ [µm]	σ_X	σ_Y	σ_Z	σ_x	σ_y	μ_X	μ_Y	μ_Z	μ_x	μ_y
1	4	9	3	72	581	0.82	0.91	0.82	1.23	0.7	0.4	0.59	0.58	0.85	1.5	1.5
2	4	9	8	67	598	0.84	0.29	0.27	0.62	0.7	0.5	0.39	0.31	0.75	0.9	0.7

Ver ....Version
Im ....Number of images
AP ....Number of additional parameters
Co ....Number of control points
Ch ....Number of check points
r ....Redundancy

$\hat{\sigma}_0$ ....Standard deviation of measured image coordinates a posteriori
σ_{XYZ} ....Theoretical precision in object space
σ_{xy} ....Theoretical precision in image space
μ_{XYZ} ....Empirical accuracy in object space
μ_{xy} ....Empirical accuracy in image space

[1]Values obtained after a spatial similarity transformation of all check point coordinates onto their reference values.

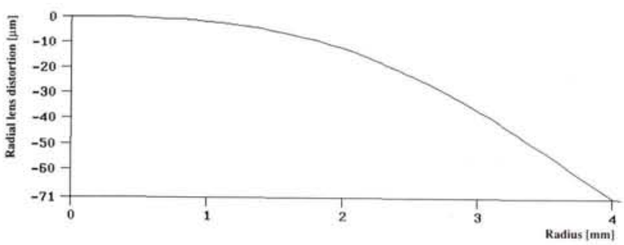


Figure 5. Radial lens distortion of the JVC video camera.

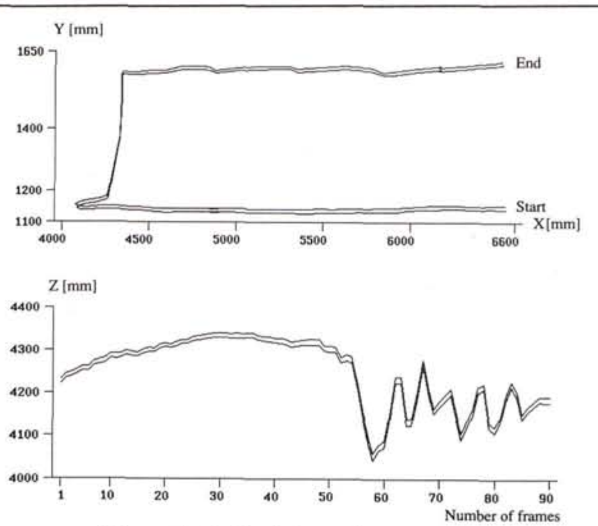


Figure 6. Path of the video camera.

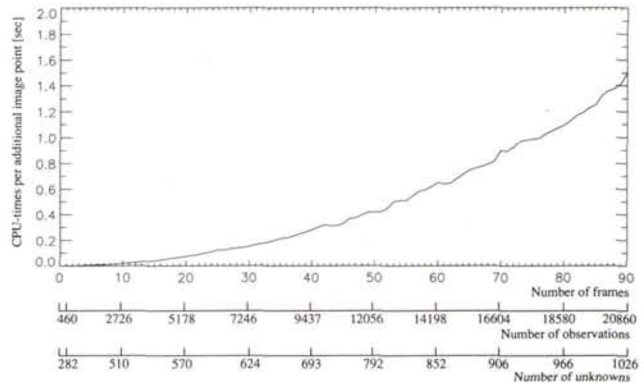


Figure 7. CPU-times for the inclusion of one additional image point into the sequence (sequential estimation in OLTRIS on a SPARCstation 1+, Sun Microsystems).

−11 mm in the y-direction, and 12 mm in the z-direction. This difference can be attributed to the absence of systematic error compensation in the sequentially estimated version.

The important comparison to be made here between the two adjustment techniques, simultaneous and sequential, relates to their respective computation times (CPU) for updating the normal equation system and calculating the solution vector. In OLTRIS, it is possible to perform sequential update with Givens transformations and simultaneous adjustment with Cholesky factorization and back-substitution. Computing times (CPU) for the updating of the solution vector when including the observations of one additional image point are illustrated in Figure 7. The plotted line shows the increase of CPU time consumption, depending on the number of frames, observations, and unknowns, respectively. The computation time measured was between 0.01 seconds per additional image point measurement at the start phase of the triangulation and 1.54 seconds at the stage of the last frame of the sequence. In comparison with these results, this speed could not be achieved with a simultaneous adjustment. Here, the normal equation system is relinearized, if the updating of the solution vector is requested after adding an image point measurement into the normals. For that, forming and solving the normal equations during the triangulation takes 3 seconds per iteration at the stage of ten introduced image frames, including 2726 observations, and approximately 20 seconds at the stage of 40 frames (9437 observations). This is approximately by a factor 70 worse than the sequential mode and is far away from video real-time.

Conclusions

Our investigations have shown that sequential estimation in a general point positioning and camera orientation module (bundle adjustment) using Givens transformations can result in very short response times for system updating. In our example, the insertion of one additional image point required

0.01 seconds at the stage of the first CCD frame and 1.54 seconds at 88 frames on a Sun SPARCstation 1+. The simultaneous solution required by a factor 70 higher computing times. Thus, within this computer environment, an image point insertion (and deletion) at video rate (0.02 sec) can be achieved at a system size of ten frames. This excellent computational performance makes the procedure of sequential updating of bundle systems by Givens transformations particularly useful in time-constrained machine vision and robot vision applications. From a system point of view, however, object space feature positioning may be only a minor portion of the overall computing time budget. Because image analysis and image understanding operations can easily chew up a large amount of computing time, it should be worthwhile to investigate possible sequential formulations of related algorithms.

As a by-product of our investigations we could show that, even with an "amateur" TV video camera with an integrated analog storage device, a fairly good accuracy (1/10th of a pixel from 3D check points) can be achieved. We believe that even better accuracies are possible if emphasis is put on a more sophisticated procedure for systematic error compensation. This opens interesting perspectives for the use of TV video cameras in a great variety of measurement applications.

Acknowledgment

The software package OLTRIS was developed in the project "On-line point positioning with single frame camera data," which was sponsored by the U.S. Government, Department of Defense, represented through its European Research Office of the U.S. Army in London. This support is gratefully acknowledged.

References

Baltsavias, E., A. Gruen, and M. Meister, 1991. DOW - A System for Generation of Digital Orthophotos from Aerial and SPOT Images, presented paper, ACSM/ASPRS Annual Convention, 25–29 March, Baltimore, Maryland.

Beyer, H.A., 1987. *Einige grundlegende Designfragen für die Entwicklungsumgebung für Digitale Photogrammetrie auf den Sun Workstations (DEDIP, Development Environment for Digital Photogrammetry)*, internal report, ETH-Zurich, Switzerland.

———, 1992. *Geometric and Radiometric Analysis of a CCD-Camera Based Photogrammetric Close-Range System*, Dissertation No. 9701, ETH-Zurich.

Beyer, H., Th. Kersten, and A. Streilein, 1992. Metric Accuracy Performance of Solid-State Camera Systems, paper presented at the SPIE OE/Technology '92 — Videometrics conference, 15–20 November, 1992, Boston, Massachusetts. Published in SPIE Vol. 1820, *Videometrics*.

Blais, J.A.R., 1983. Linear Least-Squares Computations Using Givens Transformations, *The Canadian Surveyor*, 37(4):225–233.

Boullion, T., and P. Odell, 1971. *Generalized Inverse Matrices*, Wiley Interscience, Wiley and Sons, New York.

Deriche, R., and O. Faugeras, 1990. *Tracking Line Segments*, Proceedings of the First European Conference of Computer Vision, Antibes, France, April, pp. 259–268.

Edmundson, K.L., 1991. *On-Line Triangulation of Sequential Stereo-Pairs Using Givens Transformations without Square Roots*, Master Thesis, Department of Geodetic Sciences and Surveying, The Ohio State University, Columbus, Ohio.

Gentleman, W.M., 1973. Least Squares Computation without Square Roots, *Journal of the Institute of Mathematical Applications*, (12):329–336.

Golub, G.H., 1969. Matrix Decompositions and Statistical Calculations, *Statistical Calculations*, (R.C. Milton and J.A. Nelder, editors) Academic Press, New York, pp. 365–397.

Gruen, A., 1982. An Optimum Algorithm for On-Line Triangulation, *Int. Arch. of Photogrammetry*, Vol. 24-III, Helsinki, Finland.

———, 1985a. Algorithmic Aspects in On-Line Triangulation, *Photogrammetric Engineering & Remote Sensing*, 51(4):419–436.

———, 1985b. Data Processing Methods for Amateur Photographs, *Photogrammetric Record*, 11(65):567–579.

Gruen, A., and H.A. Beyer, 1990. DIPS II — Turning a Standard Computer Workstation into a Digital Photogrammetric Station, *Int. Arch. of Photogrammetry*, 28(Part II):247–255, and *Zeitschrift für Photogrammetrie und Fernerkundung*, 1991(1):2–10.

Holm, K.R., 1989. Test of Algorithms for Sequential Adjustment in On-Line Phototriangulation, *Photogrammetria*, 43:143–156.

Kersten, Th., A. Gruen, and K.R. Holm, 1992. *On-Line Point Positioning with Single Frame Camera Data*, DoD Final Technical Report No. 7, Swiss Federal Institute of Technology, IGP-Bericht, No. 197, ETH-Zurich, Switzerland.

Lawson, Ch. L., and R.J. Hanson, 1974. *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, N.J.

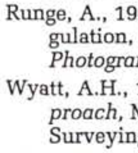
Matthies, L.M., R. Szeliski, and T. Kanade, 1989. Kalman Filter-based Algorithms for Estimating Depth from Image Sequences, *Int. Journal of Computer Vision*, 3(3):209–238.

Runge, A., 1987. The Use of Givens Transformation in On-Line Triangulation, *Proceedings, ISPRS Conference on "Fast Processing of Photogrammetric Data"*, Interlaken, Switzerland, 2–4 June.

Wyatt, A.H., 1982. *On-Line Triangulation — An Algorithmic Approach*, Master Thesis, Department of Geodetic Sciences and Surveying, The Ohio State University, Columbus, Ohio.

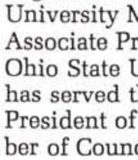
Zhang, Z., and O. Faugeras, 1990. Motion Tracking in a Sequence of Stereo Frames, *Proceedings, 9th European Conference of Artificial Intelligence*, Stockholm, Sweden, August, pp. 747–752.

(Received 17 February 1993; accepted 3 August 1993)



Armin Gruen

Armin Gruen is Professor for Photogrammetry (since 1984) and currently Head of the Institute of Geodesy and Photogrammetry, ETH Zürich, Switzerland. He received the Dipl.-Ing. degree in Geodetic Science in 1968 and the doctorate degree in Photogrammetry in 1974, both from the Technical University Munich, Germany. He has held an appointment as Associate Professor at the Department of Geodetic Science, the Ohio State University, Columbus, Ohio (1981-1984). Dr. Gruen has served the ISPRS as Working Group Chairman (1984-88), President of Commission V (1988-92), and is currently Member of Council as Second Vice President (1992-96).



Thomas P. Kersten

Thomas Kersten received the Dipl.-Ing. degree in Geodetic Science in 1988 from the University of Hannover, Germany. Since 1989, he has been a research and teaching assistant at the Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology, ETH Zurich. From 1989 to 1991, Mr. Kersten did research in on-line triangulation for photogrammetric point positioning in aerial applications. His current research interests include, among others, the use of on-line triangulation algorithms in robot vision applications.