CONTINUOUS SELF-CALIBRATION AND EGO-MOTION DETERMINATION OF A MOVING CAMERA BY OBSERVING A PLANE

Jochen Meidow, Michael Kirchhof

FGAN-FOM Research Institute for Optronics and Pattern Recognition,
Gutleuthausstr. 1, 76275 Ettlingen, Germany
{meidow|kirchhof}@fom.fgan.de

KEY WORDS: camera calibration, homography, parameter estimation, Kalman filter

ABSTRACT:

In many vision applications the depth relief of an observed scene is small compared with the extent of the image. Beside man-made environments these scenes may be approximated by a plane. Due to environmental influences camera parameters can gradually change which motivates the need for a continuous self-calibration. Based on the theory of recursive parameter estimation we present an update scheme for the parameters to handle endless video streams in real time. The geometric parametrization of the frame to frame homographies allows to incorporate information of other sensors. The application is the visual navigation of robots and unmanned aerial vehicles moving above nearly planar environments. The approach will be empirically evaluated with a synthetic data set and demonstrated with a real data set. A typical example is given by a real data set of an indoor robot observing the ground floor with one camera.

1 INTRODUCTION

Motivation. Many vision applications deal with a small depth relief of an observed scene compared with the extent of the image. Often these scenes can be approximated by a plane e.g. the ground plane. Especially man-made environments essentially consist of planes. Therefore a homography based approach is proper to model such configurations. The main application at our focus is an indoor robot observing the ground floor in order to determine its trajectory visually. The odometry of an indoor robot is often unreliable due to the slip of the wheels. Therefore image based determination of ego-motion should be used in addition to standard odometry.

Due to environmental influences camera parameters can gradually change. Therefore the investigation of odometry determined from a monocular sensor observing a ground plane requires continuous self-calibration. The application also requires to handle an endless video stream in real time. Therefore a method apart from the computational effort of bundle adjustment is needed.

Related Work. Fundamental work on geometric decomposition of homography was done in (Faugeras and Lustman, 1988). But this method requires a proper calibration of the used cameras. Camera self-calibration from views of a 3D scene has widely investigated, see for example (Pollefeys et al., 1999, Maybank and Faugeras, 1992). But it is known that these techniques in general fail for planar or almost planar scenes since they run into singularities. Nevertheless, self-calibration from planar scenes is possible (Triggs, 1998, Zhang, 1999, Malis and Cipolla, 2002).

In (Zhang, 1999) every available metric information of the scene is used and the calibration is determined by plane to frame homographies. A closed form solution is given and improved by non-linear optimization based on maximum likelihood criterion. The work was inspired by (Triggs, 1998). He computes constraints from the dual image of the absolute conic again from plane to frame homographies. Frame to frame homographies are used in (Malis and Cipolla, 2002) to build up a so-called super-collineation-matrix and to enforce multi-view constraints e.g. rank constraint. The cost-function for the self-calibration is then given by the difference of the eigenvalues of matrices with similar properties to the essential matrix. Fundamental work on relative orientation was done by Nistér. His solution to the five-point relative pose problem (Nistér, 2004) using the essential matrix can deal 3D as well as planar scenes. But the determination of the essential matrix requires calibrated cameras and the precision of the relative orientation is decreased when the scene becomes planar. Therefore the relative orientation computed from the essential matrix is preferable if the cameras move with known calibration in 3D environments.

Contribution. We present odometry visually determined from a monocular camera including update of the intrinsic camera parameters. We have to deal with endless video streams and have to ensure real time capability. The update of the intrinsic camera parameters is relevant because for example the focal length varies depending on the temperature. Nevertheless we can assume an initial guess because the used cameras are known.

It is a well known result in photogrammetry that the best results can be obtained by bundle adjustment (McGlone et al., 2004). On the opposite the relative orientation between subsequent frames can be computed very efficiently. One possible compromise is an incremental bundle adjustment. A typical method for an incremental bundle adjustment of homographies can be found in (Zelnik-Manor and Irani, 2002, Han and Kanade, 1998).

We introduce a different technique inspired by Kalman filtering (Welch and Bishop, 2003, Kalman, 1960), where the information of the past is subsumed in one parameter vector and its covariance matrix. The method reflects the possible smooth change of the calibration and also allows to incorporate other sensor information (e. g. GPS, INS, odometry) by using a geometric parametrization. The computational effort is comparable to adjustment over two homographies and is independent on the number of frames taken into account for estimating the current state.

Notation. For formulation and representation we use the framework of algebraic projective geometry. Homogeneous vectors and matrices will be denoted with upright boldface letters, e.g. $\mathbf{x}$ or $\mathbf{H}$, Euclidean vectors or matrices with slanted boldface letters, e.g. $\boldsymbol{x}$ or $\boldsymbol{H}$. In homogeneous coordinates '=' means an assignment or an equivalence up to an scaling factor $\lambda \neq 0$. Many

parameters have to be represented in various coordinate systems. Observations in the coordinate system S_k attached to the k-th frame are denoted by an upper index e.g. $\overset{k}{\boldsymbol{x}}$. Relative orientations or mappings between two cameras are written as $(\boldsymbol{R}_{kl}, \boldsymbol{t}_{kl})$ representing the motion from S_k to S_l.

2 MODELING

2.1 Basic Relations

For a 3D point $\boldsymbol{X}_i$ in the plane π the incidence relation

$$\boldsymbol{n}_k^\mathsf{T} \overset{k}{\boldsymbol{X}}_i - d_k = 0 \tag{1}$$

holds, where the plane is represented by its normal vector $\boldsymbol{n}_k$ and its distance d_k to the origin of the camera coordinate system S_k. The normal vector $\boldsymbol{n} = [n_X, n_Y, n_Z]^\mathsf{T}$ is oriented so that it points towards the camera, i.e. $n_Z \leq 0$. The representation of the point $\boldsymbol{X}_i$ in an other camera coordinate system S_l is equivalent to the coordinate transformation

$$\overset{k}{\boldsymbol{X}}_i = \boldsymbol{R}_{kl} \overset{l}{\boldsymbol{X}}_i - \boldsymbol{t}_{kl} \tag{2}$$

being a rigid 3D motion. Equations (1) and (2) reveal that the object points and their corresponding image points are related by the 2D homography

$$\overset{k}{\boldsymbol{X}}_i = \left(\boldsymbol{R}_{kl} - \boldsymbol{t}_{kl}\boldsymbol{n}_l^\mathsf{T}/d_l\right) \overset{l}{\boldsymbol{X}}_i = \boldsymbol{H}_{lk} \overset{l}{\boldsymbol{X}}_i \tag{3}$$

induced by the plane π. The rotation $\boldsymbol{R}_{kl}$ and the translation $\boldsymbol{t}_{kl}$ constitute the relative orientation between the two cameras with five parameters since only the ratio $\boldsymbol{t}_{kl}/d_l$ is determinable from (3), cf. figure 1. The decomposition of $\boldsymbol{H}_{lk}$ according to (3) has up to eight solutions which can be reduced to two reasonable solutions (Faugeras and Lustman, 1988).

Assuming straight-line preserving cameras the projection of the object points $\boldsymbol{X}_i$ is $\mathbf{x}_{ik} = \mathbf{K}_k \overset{k}{\mathbf{X}}_i$ with the homogeneous calibration matrix $\mathbf{K}_k$ introducing five additional intrinsic parameters per camera. Thus the corresponding 2D homography reads

$$\overset{k}{\mathbf{x}}_i = \mathbf{K}_k \boldsymbol{H}_{lk} \mathbf{K}_l^{-1} \overset{l}{\mathbf{x}}_i = \mathbf{H}'_{lk} \overset{l}{\mathbf{x}}_i \tag{4}$$

allowing to determine eight parameters from the correspondencies of one image pair. Note that the matrix $\boldsymbol{H}_{lk}$ — in contrast to $\mathbf{H}'_{lk}$ — can be decomposed into Euclidean entities.

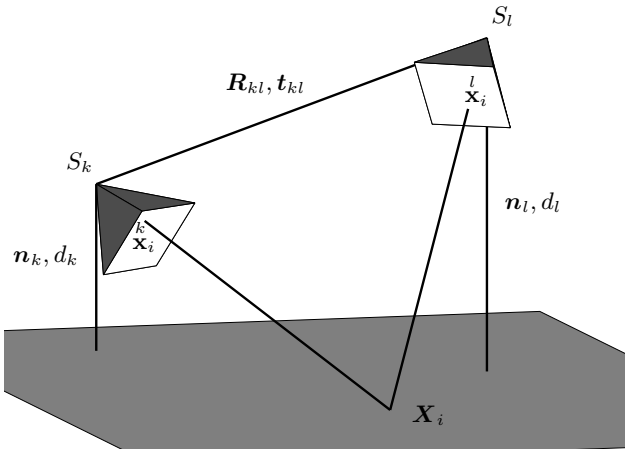

Figure 1: shows two cameras in general position and orientation observing a point $\boldsymbol{X}_i$ on the plane.

2.2 Adjustment Model

Functional Model. Without loss of generality we consider a parameter estimation for two groups of observations $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ being explicit functions of the unknown parameters $\boldsymbol{x}$, i.e. $\boldsymbol{l}_1 = \boldsymbol{f}_1(\boldsymbol{x})$ and $\boldsymbol{l}_2 = \boldsymbol{f}_2(\boldsymbol{x})$. This model is the basis for recursive and sequential parameter estimation schemes, where observations are added at later stage resulting in an update for the parameters.

With the additional restrictions $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$ for the parameters the corresponding linear model for the estimated entities reads (McGlone et al., 2004)

$$\boldsymbol{\Delta l}_1 + \widehat{\boldsymbol{v}}_1 = \boldsymbol{A}_1 \widehat{\boldsymbol{\Delta x}} \tag{5}$$
$$\boldsymbol{\Delta l}_2 + \widehat{\boldsymbol{v}}_2 = \boldsymbol{A}_2 \widehat{\boldsymbol{\Delta x}} \tag{6}$$
$$\boldsymbol{H} \widehat{\boldsymbol{\Delta x}} = \boldsymbol{c} \tag{7}$$

with the Jacobians $\boldsymbol{A}$ and $\boldsymbol{H}$, the estimated corrections $\widehat{\boldsymbol{v}}$, the differences $\boldsymbol{\Delta l} = \boldsymbol{l} - \boldsymbol{f}(\boldsymbol{x}_0)$ and $\widehat{\boldsymbol{\Delta x}} = \widehat{\boldsymbol{x}} - \boldsymbol{x}_0$, and the contradictions $\boldsymbol{c} = -\boldsymbol{h}(\boldsymbol{x}_0)$, evaluated at the approximate values $\boldsymbol{x}_0$.

Stochastic Model. We assume statistically independent observation groups, i. e. $\boldsymbol{\Sigma}_{l_1 l_2} = \boldsymbol{O}$, with the known covariance matrices $\boldsymbol{\Sigma}_{l_1 l_1}$ and $\boldsymbol{\Sigma}_{l_2 l_2}$. These covariance matrices are related to the true covariance matrices by an unknown variance factor σ_0^2 which can be estimated from the estimated corrections $\widehat{\boldsymbol{v}}$, see below. If the initial covariance reflects correctly the uncertainty of the observations, this variance factor is $\sigma_0^2 = 1$ (McGlone et al., 2004).

Normal Equations. Minimizing the squared and weighted sum of residuals

$$\Omega = \widehat{\boldsymbol{v}}_1^\mathsf{T} \boldsymbol{\Sigma}_{l_1 l_1}^{-1} \widehat{\boldsymbol{v}}_1 + \widehat{\boldsymbol{v}}_2^\mathsf{T} \boldsymbol{\Sigma}_{l_2 l_2}^{-1} \widehat{\boldsymbol{v}}_2 \tag{8}$$

under the linear constraints (7) the corresponding normal equation system becomes

$$\begin{bmatrix} \boldsymbol{N} & \boldsymbol{H}^\mathsf{T} \\ \boldsymbol{H} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\Delta x}} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \boldsymbol{h} \\ \boldsymbol{c} \end{bmatrix} \tag{9}$$

with the Lagrangian multipliers $\boldsymbol{\lambda}$ and

$$\boldsymbol{N} = \boldsymbol{A}_1^\mathsf{T} \boldsymbol{\Sigma}_{l_1 l_1}^{-1} \boldsymbol{A}_1 + \boldsymbol{A}_2^\mathsf{T} \boldsymbol{\Sigma}_{l_2 l_2}^{-1} \boldsymbol{A}_2 \tag{10}$$
$$\boldsymbol{h} = \boldsymbol{A}_1^\mathsf{T} \boldsymbol{\Sigma}_{l_1 l_1}^{-1} \boldsymbol{\Delta l}_1 + \boldsymbol{A}_2^\mathsf{T} \boldsymbol{\Sigma}_{l_2 l_2}^{-1} \boldsymbol{\Delta l}_2. \tag{11}$$

The solution of the system (9) is identical to the solution of the alternative normal equation system (McGlone et al., 2004)

$$\begin{bmatrix} \boldsymbol{A}_1^\mathsf{T} \boldsymbol{\Sigma}_{l_1 l_1}^{-1} \boldsymbol{A}_1 & \boldsymbol{A}_2^\mathsf{T} & \boldsymbol{H}^\mathsf{T} \\ \boldsymbol{A}_2 & -\boldsymbol{\Sigma}_{l_2 l_2} & \boldsymbol{O}^\mathsf{T} \\ \boldsymbol{H} & \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\Delta x}} \\ \boldsymbol{\mu} \\ \boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1^\mathsf{T} \boldsymbol{\Sigma}_{ll} \boldsymbol{\Delta l}_1 \\ \boldsymbol{\Delta l}_2 \\ \boldsymbol{c} \end{bmatrix} \tag{12}$$

with the additional Lagrangian multipliers $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. This solution can address the situation $\boldsymbol{\Sigma}_{l_2 l_2} = \boldsymbol{O}$, that is one can fix parameters by setting their covariances to zero. Furthermore one can undo information by substituting $-\boldsymbol{\Sigma}_{l_2 l_2}$ with $\boldsymbol{\Sigma}_{l_2 l_2}$.

By applying the definition for a pseudo inverse, i.e. $\boldsymbol{\Sigma}^+ = \boldsymbol{\Sigma}^+ \boldsymbol{\Sigma} \boldsymbol{\Sigma}^+$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma} \boldsymbol{\Sigma}^+ \boldsymbol{\Sigma}$, it can be shown that the equivalence of (9) and (12) holds for singular covariance matrices $\boldsymbol{\Sigma}_{l_2 l_2}$, too. Thus the weighted sum

$$\Omega = \widehat{\boldsymbol{v}}_1^\mathsf{T} \boldsymbol{\Sigma}_{l_1 l_1}^{-1} \widehat{\boldsymbol{v}}_1 + \widehat{\boldsymbol{v}}_2^\mathsf{T} \boldsymbol{\Sigma}_{l_2 l_2}^{+} \widehat{\boldsymbol{v}}_2 \tag{13}$$

becomes minimal subject to the constraints for the parameters.

Precision. The covariance matrix of the estimated parameters $\widehat{\boldsymbol{x}}$ results from the inverse normal equation matrices by

$$\begin{bmatrix} \boldsymbol{\Sigma}_{\widehat{x}\widehat{x}} & . \\ . & . \end{bmatrix} = \begin{bmatrix} \boldsymbol{N} & \boldsymbol{H}^\mathsf{T} \\ \boldsymbol{H} & \boldsymbol{O} \end{bmatrix}^{-1} \tag{14}$$

or

$$\begin{bmatrix} \boldsymbol{\Sigma}_{\widehat{x}\widehat{x}} & . & . \\ . & . & . \\ . & . & . \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1^\mathsf{T}\boldsymbol{\Sigma}_{l_1l_1}^{-1}\boldsymbol{A}_1 & \boldsymbol{A}_2^\mathsf{T} & \boldsymbol{H}^\mathsf{T} \\ \boldsymbol{A}_2 & -\boldsymbol{\Sigma}_{l_2l_2} & \boldsymbol{O}^\mathsf{T} \\ \boldsymbol{H} & \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}^{-1} \tag{15}$$

and reveals the theoretical precision of the parameters.

The unknown variance factor can be estimated by

$$\widehat{\sigma}_0^2 = \frac{\Omega}{R} \tag{16}$$

with the redundancy R of the system. Thus the empirical precision becomes

$$\widehat{\boldsymbol{\Sigma}}_{\widehat{x}\widehat{x}} = \widehat{\sigma}_0^2\boldsymbol{\Sigma}_{\widehat{x}\widehat{x}}. \tag{17}$$

2.3 Prior Information

The need for introducing prior information into the estimation process is manifold: Geometric weak configurations of the sensors or the objects may lead to degeneracies resulting in multiple solutions, a parametric family of solution, or no solution at all. To cope with these degeneracies prior information can be introduced in a Bayesian manner by parameter values and their uncertainties in form of covariance matrices. These can be either real observations of the sought parameters or fictitious observations. Beside the eventually necessity to effect determinability the use of prior information leads to stabilization and smoothness.

Applying (9) in a rigorous successive way would mean keep the entire information of the past for the estimations of future states. But a sneaking change of parameters due to changing environmental conditions requires a changing point of linearization, too. To cope with this situation we introduce a factor $\alpha \in (0, 1)$ which reflects the amount of information which should be used to estimate the current state of the parameters (memory length). Furthermore this factor determines the relative weight for the contributions of current data and the prior, i.e. the past information.

The amount of data taken from the past to determine the current state can be estimated in advance. If the number m of observations per image pair is constant the limit of the geometric series $s = \sum_{i=1}^{n} m\alpha^{i-1}$ approaches $s = 1/(1-\alpha)$ for $n \to \infty$ and $\alpha < 1$. Thus for $\alpha = 0.95$ the amount of approximate 20 image pairs will be used to estimated the current state for instance.

2.4 Relation to Kalman filtering

Neglecting the constraints (7) in (9) yields the formulas for the recursive parameter estimation or the Kalman filtering (Kalman, 1960). With the estimation $\widehat{\boldsymbol{\Delta x}}_1 = \boldsymbol{\Sigma}_{\widehat{x}_1\widehat{x}_1}\boldsymbol{A}_1\boldsymbol{\Sigma}_{l_1l_1}^{-1}\boldsymbol{\Delta l}_1$ because of the first observation group $\boldsymbol{l}_1$ and the corresponding covariance matrix $\boldsymbol{\Sigma}_{\widehat{x}_1\widehat{x}_1} = (\boldsymbol{A}_1\boldsymbol{\Sigma}_{l_1l_1}^{-1}\boldsymbol{A}_1)^{-1}$ the Kalman-Filter gain matrix is (Welch and Bishop, 2003)

$$\boldsymbol{F} = \boldsymbol{\Sigma}_{\widehat{x}_1\widehat{x}_1}\boldsymbol{A}_2^\mathsf{T}\left(\boldsymbol{\Sigma}_{l_2l_2} + \boldsymbol{A}_2\boldsymbol{\Sigma}_{\widehat{x}_1\widehat{x}_1}\boldsymbol{A}_2^\mathsf{T}\right)^{-1}. \tag{18}$$

The update of the parameters due to the new observations $\boldsymbol{l}_2$ is:

$$\widehat{\boldsymbol{\Delta x}} = \widehat{\boldsymbol{\Delta x}}_1 + \boldsymbol{F}(\boldsymbol{A}_2\widehat{\boldsymbol{\Delta x}}_1 - \boldsymbol{\Delta l}_2) \tag{19}$$

$$\boldsymbol{\Sigma}_{\widehat{x}\widehat{x}} = (\boldsymbol{I} - \boldsymbol{F}\boldsymbol{A}_2)\boldsymbol{\Sigma}_{\widehat{x}_1\widehat{x}_1} \tag{20}$$

These equations constitute the so-called measurement update (correction) of the extended Kalman filter.

In an intermediate step the time update equations (prediction) are responsible for forward projection of the parameters, e.g. because of a change in the coordinate system accompanied by error propagation.

3 REALIZATION

3.1 Parametrization

For the recursive parameter estimation we set the distances $d_k \doteq 1$ for all pairs of images. For the plane normal vectors $\boldsymbol{n}_k$ we assume all three components to be unknown and introduce the (hard) constraint

$$\boldsymbol{n}_k^\mathsf{T}\boldsymbol{n}_k = 1. \tag{21}$$

The over-parametrization is justified by the fact that eq. (21) is bilinear in the parameters and therefore convenient for linearization. For the rotation $\boldsymbol{R}_{kl}$ three parameters $\boldsymbol{r}_{kl}$ are introduced with $l = k + 1$.

For the single camera we choose the affine model

$$\mathbf{K} = \begin{bmatrix} c & cs & x_0 \\ 0 & cm & y_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{22}$$

with the camera constant c, the scale m, the skew s, and the principal point $[x_0, y_0]$. The intrinsic camera parameters are therefore $\boldsymbol{k} = [c, m, s, x_0, y_0]^\mathsf{T}$. Parameters with known or assumed values can easily be fixed by setting the corresponding variances in $\boldsymbol{\Sigma}_{l_2l_2}$ to zero.

Dropping the indices the parameter vector is $\boldsymbol{x} = [\boldsymbol{k}^\mathsf{T}, \boldsymbol{n}^\mathsf{T}, \boldsymbol{r}^\mathsf{T}, \boldsymbol{t}^\mathsf{T}]^\mathsf{T}$, where $\boldsymbol{k}$ are global parameters, $\boldsymbol{n}$ are global parameters which have to be propagated, and $\boldsymbol{r}$ and $\boldsymbol{t}$ are local parameters for each consecutive image pair.

3.2 Observations, Weights and Model Validation

The correspondencies of image points can be established essentially in two ways: The first method first extracts interest points in both images and then finds the correspondencies, e.g. by correlation followed by RANSAC (Fischler and Bolles, 1981). The second methods extracts interest points in one image and searches them in the second image, e.g. by a tracker (Lucas and Kanade, 1981).

Empirical covariance matrices of the sub-pixel positions can be derived anyway – either by an residual-based or an derivative-based approach (Kanazawa and Kanatani, 2001). The inverse covariance matrices are then the weight matrices for the adjustment procedure. The first observation group $\boldsymbol{l}_1$ consists of the coordinates of the set of observed points. By considering errors in one image for each correspondence i and mapping $\boldsymbol{H}$

$$\overset{k}{x}_i + \overset{k}{v}_{x_i} = \frac{H_{11}\overset{l}{x}_i + H_{12}\overset{l}{y}_i + H_{13}}{H_{31}\overset{l}{x}_i + H_{32}\overset{l}{y}_i + H_{33}} \tag{23}$$

$$\overset{k}{y}_i + \overset{k}{v}_{y_i} = \frac{H_{21}\overset{l}{x}_i + H_{22}\overset{l}{y}_i + H_{23}}{H_{31}\overset{l}{x}_i + H_{32}\overset{l}{y}_i + H_{33}} \tag{24}$$

holds with the reprojection errors v_i.

The application of the RANSAC procedure enforces the validity of the mapping model (4). Thus model violations by non-planar

scenes can be treated as long as a dominant plane is visible. Furthermore, with the realistic weights for the observations the estimation of the variance factor (16) can be tested statistically since its expectation value is one.

3.3 Approximation Values

In general the parameter estimation model requires approximation values $\boldsymbol{x}_0$ for the sought parameters. For applications processing a dense video stream it is usually sufficient to use the estimation results from the previous image pair as approximation values for the next image pair. Furthermore for the relative rotation $\boldsymbol{R}_0 = \boldsymbol{I}_3$ holds. For applications with a long base line approximation values can be obtained from the decomposition (3) with an approximately given calibration matrix.

3.4 Application I: Calibration of an Airborne Camera

For a high-flying airborne camera the observed scenes usually appear flat. The heights above ground and the relative orientation between two consecutive frames is individual. Thus we have 3 rotation parameters and 3 translation components for each frame pair and quasi global parameters $\boldsymbol{k}$ and $\boldsymbol{n}$. Prior information is introduced by the estimated parameters and their uncertainties from the respective previous image pair. The normal vector has to be transformed into the next camera coordinate system. Therefore the prior information reads $\boldsymbol{l}_2 = (\boldsymbol{k}_l^\mathsf{T}, \boldsymbol{n}_l^\mathsf{T})^\mathsf{T}$ with

$$\begin{aligned} \boldsymbol{k}_k &= \boldsymbol{k}_l & (25)\\ \boldsymbol{n}_l &= \boldsymbol{R}_{lk}\boldsymbol{n}_k, & (26) \end{aligned}$$

$\boldsymbol{R}_{lk} = \boldsymbol{R}(\boldsymbol{r}_{lk})$, and the covariance matrix $\boldsymbol{\Sigma}_{l_2l_2}$ estimated with (14) or (15) from the previous stage. The Jacobian of (25) and (26) is

$$\boldsymbol{A}_2 = \begin{bmatrix} \boldsymbol{I}_5 & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{R}_{lk}^\mathsf{T} & (\partial \boldsymbol{R}_{lk}^\mathsf{T}/\partial \boldsymbol{r}_{lk})\boldsymbol{n}_k & \boldsymbol{O} \end{bmatrix}. \quad (27)$$

Note that $\boldsymbol{\Sigma}_{l_2l_2}$ has a rank deficiency of 1 due to the constraint $\boldsymbol{n}_k^\mathsf{T}\boldsymbol{n}_k = 1$ on the parameters. Its null space $\mathcal{N}(\boldsymbol{\Sigma}_{l_2l_2}) = (\boldsymbol{0}^\mathsf{T}, \boldsymbol{n}_k^\mathsf{T})^\mathsf{T}$ can be use to compute the pseudo inverse $\boldsymbol{\Sigma}_{l_2l_2}^+$ if needed.

For a metric reconstruction of the sensor trajectory the scale has to be introduced and updated by $d_k = d_j + \boldsymbol{n}_j^\mathsf{T}\boldsymbol{t}_{jk}$ (Han and Kanade, 1998).

3.5 Application II: Camera Calibration for Indoor Robots

For an indoor robot moving on a ground plane the camera height above ground $d_l = d_k = d$ is constant. Furthermore the rotation is restricted to rotations by an angle ϕ around the plane normal $\boldsymbol{n}$ being the only rotation parameter $\boldsymbol{r} = (\phi)$. Thus an appropriate parametrization of the rotation matrix is the angle-axis-representation

$$\boldsymbol{R}(\boldsymbol{n}, \phi) = \cos(\phi)\boldsymbol{I}_3 + \sin(\phi)\boldsymbol{S}_n + (1 - \cos(\phi))\boldsymbol{D}_n \quad (28)$$

with $\boldsymbol{S}_n$ inducing cross product $\boldsymbol{S}_n\boldsymbol{m} = \boldsymbol{n} \times \boldsymbol{m}$ and $\boldsymbol{D}_n$ denoting the dyadic product $\boldsymbol{D}_n = \boldsymbol{n}\boldsymbol{n}^\mathsf{T}$. The plane normal is perpendicular to the translation vector and therefore

$$\boldsymbol{n}^\mathsf{T}\boldsymbol{t} = 0 \quad (29)$$

holds and can be used as an additional constraint.

Since the normal vector remains unaffected by the rotation, the prior information is simply $\boldsymbol{k}_l = \boldsymbol{k}_k$ and $\boldsymbol{n}_l = \boldsymbol{n}_k$ with the Jacobian

$$\boldsymbol{A}_2 = \begin{bmatrix} \boldsymbol{I}_5 & \boldsymbol{O} & \boldsymbol{0} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I}_3 & \boldsymbol{0} & \boldsymbol{O} \end{bmatrix} \quad (30)$$

w.r.t. to all parameters $\boldsymbol{x} = [\boldsymbol{k}^\mathsf{T}, \boldsymbol{n}^\mathsf{T}, \phi, \boldsymbol{t}^\mathsf{T}]^\mathsf{T}$.

3.6 Algorithm

The outline of the proposed algorithm is as follows: For each consecutive frame pair do

1. *Feature extraction*. Determine interest points in the first image, e.g. by the Förstner operator (Förstner and Gülch, 1987).
2. *Tracking*. Find the corresponding points in the subsequent image, e.g. with the KLT-tracker (Lucas and Kanade, 1981) together with estimated covariance matrices for the estimated shifts.
3. *Outlier elimination.* Apply the RANSAC procedure to determine an inlier set (Fischler and Bolles, 1981) in conjunction with minimizing algebraic distances.
4. *Recursive parameter estimation.* Calculate the parameter updates and the corresponding covariance matrices according to (9) and (14).

For applications with a wide stereo base line the interest points have to be determined independently with sub-pixel accuracy. Furthermore, in the presence of a moderate number of outliers the RANSAC procedure can be replaced by an adjustment procedure with a robust cost function.

4 EXPERIMENTAL TESTS

For the evaluation of the approach we used synthetic and real data sets. With the help of the synthetic data we show that the algorithm converges to the correct solution and produces feasible results. The applicability of the approach is shown with a real data set from a camera mounted on a dolly driving through a corridor.

For the initialization of the procedure we introduced a rough guess of the parameters and their uncertainty whereas the a rank deficiency of the covariance matrix of the normal vector has been enforced by spherical normalization accompanied by error propagation (Heuel, 2004).

4.1 Synthetic Data

Test Setup. For the evaluation of the approach we simulated the data of an indoor robot with a camera moving above a virtual chess board (see figure 2). The camera moved on a circle of constant height above the plane, the angle increment for the 200 positions were 1.8 deg.

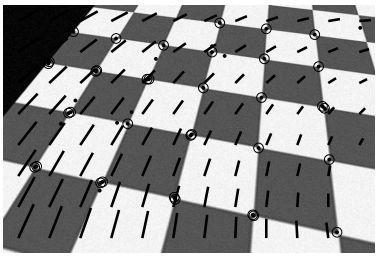

Figure 2: shows an image of the synthetic image sequence with the extracted interest points.

The image of the chess board was mapped into images of size 512 pxl × 768 pxl with the help of the calibration matrix

$$\mathbf{K} = \begin{bmatrix} 512 & 0 & 384 \\ 0 & 512 & 256 \\ 0 & 0 & 1 \end{bmatrix}. \quad (31)$$

White noise $\sigma = 2$ gr has been added to the gray values. The orientation angles of the camera w.r.t. the driving direction were

$$\begin{array}{lrl} \text{roll:} & 0.0 & \text{deg} \\ \text{nick:} & 55.0 & \text{deg} \\ \text{gear:} & 1.8 & \text{deg,} \end{array} \tag{32}$$

where the gear angle denotes the relative change in the azimuth between two consecutive camera orientations. With this test setup we have constant unknown parameters.

The images have been processed in the way described in section 3.2. Thus the results include possibly systematic errors of the tracking algorithm or the correspondences search. For the prior for the first image pair we used true parameter values (31), (32) and $\sigma_c = 20$ pxl, $\sigma_{x_0} = \sigma_{y_0} = 5$ pxl, $\sigma_{n_x} = \sigma_{n_y} = \sigma_{n_z} = 0.1$ as a rough guess of the uncertainties of the parameters $\boldsymbol{k} = (c, x_0, y_0)^\mathsf{T}$ and $\boldsymbol{n}$.

Figure 3 shows the distribution of the estimated variance factor (16). Its mean is approximately 1. Thus we conclude that the weights for the observations used within the adjustment procedure are plausible and reasonable since the model of a planar scene really holds and outliers are rejected by RANSAC.

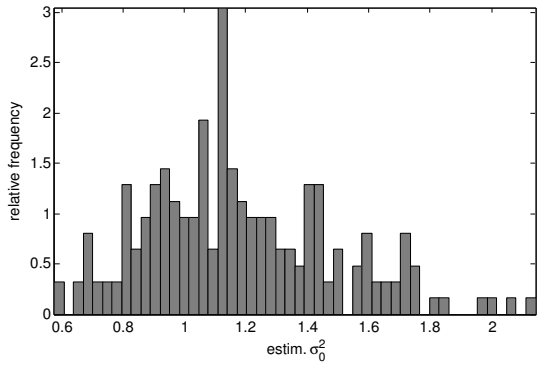

Figure 3: shows the distribution of the estimated variance factor (16) with expectation value 1.

For the evaluation the χ^2-distributed Mahalanobis distances

$$d_M = (\widehat{\boldsymbol{x}} - \widetilde{\boldsymbol{x}})^\mathsf{T} \widehat{\boldsymbol{\Sigma}}^+_{\widehat{x}\widehat{x}} (\widehat{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}) \tag{33}$$

of the estimated parameters $\widehat{\boldsymbol{x}}$ and their true values $\widetilde{\boldsymbol{x}}$ can be computed for each image pair using the pseudo inverse. Figure 4 shows the empirical distribution of these distances for all parameters $\boldsymbol{x} = [\boldsymbol{k}^\mathsf{T}, \boldsymbol{n}^\mathsf{T}, \phi, \boldsymbol{t}^\mathsf{T}]^\mathsf{T}$ and confirms the expected shape of a χ^2-distribution. For a rigorous common adjustment of all image pairs the expect average degree of freedom is 3 because the intrinsic parameters and the normal vector would be global parameters. The result illustrated in figure 4 is slightly worse because of the underlying mixed distribution.

4.2 Real Data

To demonstrate the feasibility of the approach we acquired a real data set with a video camera mounted on a dolly. The projection center was approximate 1.55 meters above the ground floor. The nick angle was approximate 33 deg, roll angle approximate 0 deg and the azimuth w.r.t. the driving direction approximately 45 deg. The image resolution is 576×720 pixel. The trajectory is curved with radii up to 2 meters. Figure 5 shows exemplary an image of the observed floor. Points on the visible part of the wall and on the fire extinguisher have been tracked, too. But since they do not lie in the dominant plane they have been discarded by the RANSAC.

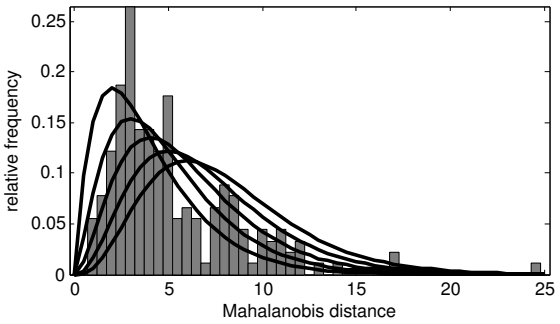

Figure 4: shows the empirical distribution of the Mahalanobis distances (33) for all parameters with the expected shape of the χ^2_4-, χ^2_5-, χ^2_6-, χ^2_7-, and χ^2_8-distribution.

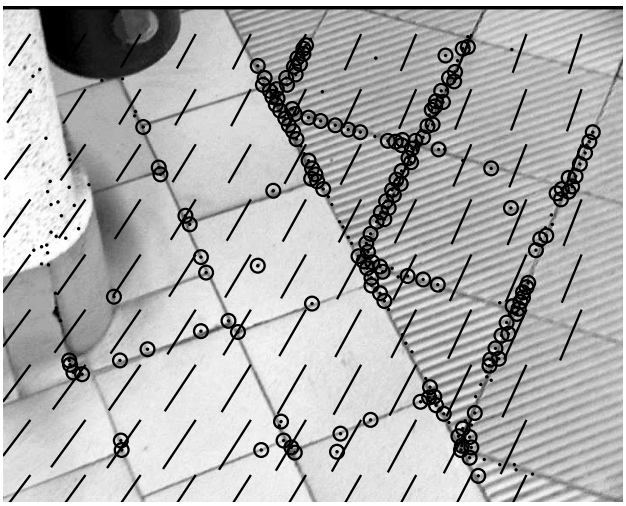

Figure 5: shows the tiled floor, the tracked interest points ($\cdot$), the points in the dominant plane determined by RANSAC ($\circ$) and the resulting motion field.

We chose a memory factor of $\lambda = 0.99$ to bridge sequence parts with critical configurations e.g. straight forward motions. Figure 6 shows the evolution of the estimated camera constant. The parameter has been initialized 50 pixels larger than the value determined separately by Zhang's calibration method (Zhang, 1999). The runs for parameters of the relative orientation are plotted in figure 7. The concatenation of these relative motions delivers the sensor trajectory.

The computations require approximately 3 seconds per image pair with a non optimized MATLAB Code including graphical output. Thus the real time capability is within easy reach.

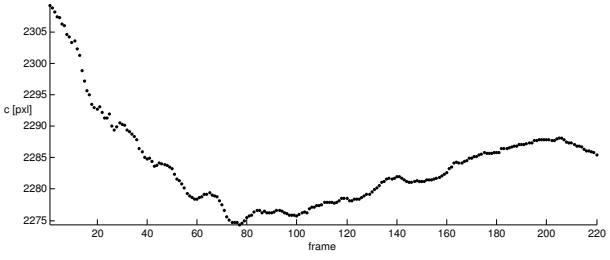

Figure 6: shows the estimated the camera constant c for the first 220 frame pairs.

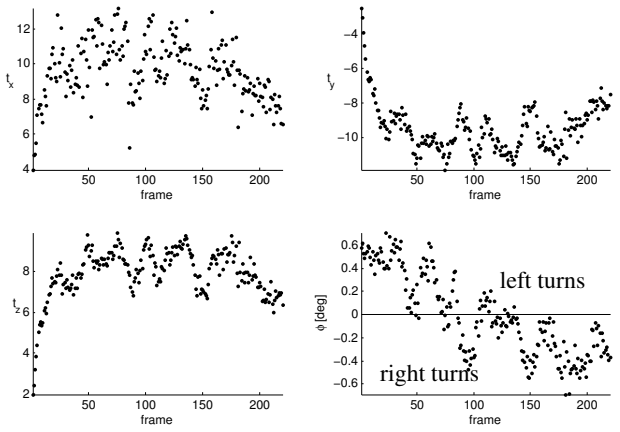

Figure 7: shows the estimated parameters t_x, t_y, t_z and ϕ of the relative orientations for the first 200 frame pairs.

5 CONCLUSIONS AND OUTLOOK

We presented and studied a continuous self-calibration method for mobile cameras observing a plane. The procedure is suitable for real time applications such as autonomously navigating indoor robots capturing endless video streams. By the application of the RANSAC the approach is able to cope with model violations as long as the scene plane dominants. Because of the geometric parametrization of the frame to frame mappings information from other sensors can easily be incorporated. Experiments with synthetic and real data sets confirm the feasibility of the approach.

Conclusions. For the determinability of the parameters a noticeable relative rotation and translation between consecutive frames is required. Image frames with almost identical orientations should be discarded e.g. by enforcing a disparity limit of say $>$10 pixel for at least one image point (Nistér, 2001). Critical motion sequences such as straight forward motions can be bridged by introducing an appropriate memory length.

Outlook. Reference data is needed to perform a more meaningful evaluation of the results for real data sets. For the intrinsic parameters this ground truth information could stem from a lab camera calibration with superior accuracy. The influence of the prior information and past information respectively can be controlled by an adaptive weighting of the information sources. Furthermore the estimation process can be stabilized by introducing further (soft) constraints which enforce a smooth trajectory of the sensor.

REFERENCES

Faugeras, O. and Lustman, F., 1988. Motion and Structure from Motion in a piecewise planar Environment. International Journal of Pattern Recognition in Artificial Intelligence 2, pp. 485–508.

Fischler, M. A. and Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the Association for Computing Machinery 24(6), pp. 381–395.

Förstner, W. and Gülch, E., 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In: ISPRS Intercommission Workshop, Interlaken.

Han, M. and Kanade, T., 1998. Homography-Based 3D Scene Analysis of Video Sequences. In: Proceedings of the DARPA Image Understanding Workshop.

Heuel, S., 2004. Uncertain Projective Geometry. Statistical Reasoning in Polyhedral Object Reconstruction. Lecture Notes in Computer Science, Vol. 3008, Springer.

Kalman, Rudolph, E., 1960. A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME–Journal of Basic Engineering 82(Series D), pp. 35–45.

Kanazawa, Y. and Kanatani, K., 2001. Do we really have to consider covariance matrices for image features? In: International Conference on Computer Vision, pp. 301–306.

Lucas, B. T. and Kanade, T., 1981. An Iterative Image Registration Technique with an Application to Stereo Vision. In: Proc. of Image Understanding Workshop, pp. 212–130.

Malis, E. and Cipolla, R., 2002. Camera self-calibration from unknown planar structures enforcing the multi-view constraints between collineations. IEEE Transactions on Pattern Analysis and Machine Intelligence 4(9), pp. 1268–1272.

Maybank, S. and Faugeras, O., 1992. A theory of self-calibration of a moving camera. International Journal of Computer Vision 8(2), pp. 123–151.

McGlone, J. C., Mikhail, E. M. and Bethel, J. (eds), 2004. Manual of Photogrammetry. 5th edn, American Society of Photogrammetry and Remote Sensing.

Nistér, D., 2001. Frame decimation for structure and motion. In: Workshop on 3D Structure from Multiple Images of Large-Scale Environments, Lecture Notes on Computer Vision, Vol. 2018, pp. 17–34.

Nistér, D., 2004. An Efficient Solution to the Five-Point relative Pose Problem. IEEE Transactions on Pattern Recognition and Machine Intelligence 26(6), pp. 756–769.

Pollefeys, M., Koch, R. and Gool, L. V., 1999. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. International Journal of Computer Vision 32(1), pp. 7–25.

Triggs, B., 1998. Autocalibration from planar scenes. In: Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany.

Welch, G. and Bishop, G., 2003. An Introduction to the Kalman Filter. Technical Report TR 95-041, Department of Computer Science, Univ. of North Carolina at Chapel Hill.

Zelnik-Manor, L. and Irani, M., 2002. Multiview Constraints on Homographies. IEEE Transactions on Pattern Recognition and Machine Intelligence 24(2), pp. 214–223.

Zhang, Z., 1999. Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. In: International Conference on Computer Vision (ICCV'99), Corfu, Greece, pp. 666–673.