METROLOGY AND MEASUREMENT SYSTEMS
VOL. XV, NUMBER 4 (2008)

THAO DANG

University of Karlsruhe
Institut für Mess- und Regelungstechnik
Karlsruhe, Germany,
thao.dang@daimler.com

AN ITERATIVE PARAMETER ESTIMATION METHOD FOR OBSERVATION MODELS WITH NONLINEAR CONSTRAINTS

This article presents a parameter estimation algorithm for observation models with nonlinear constraints. A prominent example that belongs to this category is the continuous auto-calibration of stereo cameras. Here, our knowledge of the relation between the available measurements and the desired parameters is given by a nonlinear implicit constraint equation. An estimation method derived from an Iterated Extended Kalman Filter is designed for this application. Experiments are conducted with synthetic and real data. The proposed algorithm provides very good results and is readily applicable to a wider range of applications.

Keywords: recursive estimation, Kalman filter, stereo vision, self-calibration.

1. INTRODUCTION

Recursive estimation techniques are important for many computer vision problems in which image sequences have to be processed. Especially, if the application does not allow storing all previous images or if computation time is critical, algorithms such as Kalman filters are a common choice. An example of such an application in computer vision is the auto-calibration of stereo cameras.

Auto-calibration (or self-calibration) refers to the automatic determination of stereo camera parameters such as focal lengths or relative orientations of one camera with respect to the other. It is important to note that that many stereo vision applications require relative orientation with accuracy better than 10^{-2} degrees. In standard off-line calibration approaches (e.g. [14]), these parameters are recovered by observing a well-known reference object. Auto-calibration, however, estimates camera parameters by tracking points in an arbitrary image sequence acquired by a moving stereo rig [9, 7, 16, 17]. The only input data used for self-calibration are spatial and temporal image point correspondences, i.e. the coordinates of corresponding points in the left and right

[1] Received: September 10, 2008. Revised: November 20, 2008.

image as well as the 2D displacement of image points in consecutive frames of one camera. The author believes that stereo self-calibration will be useful in automotive applications. Due to large temperature changes and mechanical vibrations, we may not assume that calibration parameters remain constant over the whole life span of a vehicle. Thus, a continuous self-calibration is required that updates the camera calibration automatically and permanently while the sensor is in use.

A difficulty with camera self-calibration (as well as with many other problems in computer vision) are observation models with nonlinear constraint equations: Assume we are given noisy input data $\hat{\mathbf{x}}$,

$$\hat{\mathbf{x}}(k) = \mathbf{x}(k) + \mathbf{e}(k), \tag{1}$$

where $\mathbf{x}(k)$ is the ideal measurement vector at time k and $\mathbf{e}(k)$ indicates additive noise components. In case of stereo self-calibration, ideal measurements $\mathbf{x}(k)$ are related to the desired camera parameters $\mathbf{z}(k)$ by a nonlinear implicit constraint of the general form

$$\mathbf{h}(\mathbf{z}(k), \mathbf{x}(k)) = 0. \tag{2}$$

Such a nonlinear problem can be handled using an Iterated Extended Kalman Filter (IEKF, e.g [12, 8]). This algorithm linearizes the constraint function $\mathbf{h}(.)$ about an operation point which is defined by our best available guess of the parameter vector and the noisy measurements $\hat{\mathbf{x}}$. a standard Kalman filter innovation step is then performed to obtain a better estimate of the parameters $\mathbf{z}(k)$, which subsequently serves as a new operation point for an additional linearization. This procedure is repeated until convergence.

The observation model as formulated by Eqs. (1) and (2) is closely related to Gauss-Helmert-Models which are well known in the photogrammetry literature. As indicated in e.g. [10], better results could be achieved by computing a corrected measurement vector and linearizing the constraint equation about both the improved state estimate and the corrected observations. In this article, we present a compact derivation of an IEKF for observations models with nonlinear constraint equations and show how to update the measurement vector. It is shown that the linearization about a better operation point can improve the estimation results significantly.

The performance of our algorithm is demonstrated on the self-calibration of a stereo rig. Please note, that the auto-calibration method used here recovers only the extrinsic parameters of the cameras, i.e. the pose of one camera with respect to the other. This article extends previous work of the same author [3].

The paper is organized as follows: Sec. 2 gives a general derivation of the IEKF for observation models with nonlinear constraints. This technique is applied in Sec. 3.1 to the partial self-calibration of a stereo camera rig. An analysis of simulation runs as well as real imagery results demonstrates the advantage of our algorithm over standard methods (Sec. 3.2). Sec. 4 summarizes our results and concludes the paper.

2. ITERATIVE PARAMETER ESTIMATION

2.1. Problem formulation

In this section, we present a compact derivation of the IEKF for general observation models with nonlinear constraint equations. As in standard Kalman filtering, the estimation process is a two-stage procedure with alternating update (Sec. 2.2) and prediction steps (Sec. 2.3). Before we describe these steps for implicit observation models, we shortly outline the general assumptions underlying our filter.

As stated before, our objective is to determine an estimate $\hat{\mathbf{z}}(k)$ of the true state vector $\mathbf{z}(k)$. In case of stereo self-calibration, $\hat{\mathbf{z}}(k)$ represents the desired camera parameters. To accomplish this task, we are given at each time instant k noisy measurements $\hat{\mathbf{x}}(k) = \mathbf{x}(k) + \mathbf{e}(k)$. We assume that $\hat{\mathbf{x}}(k)$ is perturbed by additive noise $\mathbf{e}(k)$ as defined in Eq. (1). The additive noise vector $\mathbf{e}(k)$ is considered to be a realization of Gaussian white noise process with covariance matrix $\mathbf{\Sigma}_{\mathbf{ee}}(k)$. To complete our definition of the observation model, we assume that the ideal state vector $\mathbf{z}(k)$ and the (unobservable) ideal measurement vector $\mathbf{x}(k)$ are related by the nonlinear, implicit constraint equation $\mathbf{h}(\mathrm{z}(k), \mathrm{x}(k)) = 0$ as in Eq. (2).

The system model that governs the dynamics of the state vector $\mathbf{z}(k)$ is given by the following stochastic difference equation

$$\mathbf{z}(k+1) = \mathbf{f}(\mathbf{z}(k), \mathbf{u}(k) + \mathbf{w}(k)) + \mathbf{v}(k). \tag{3}$$

For camera auto-calibration, the system model defines how the camera parameters evolve over time and how they are influenced by command signals, e.g. if the focal length is intentionally adjusted to a different value. In Eq. (3), $\mathbf{f}$ is the transition function and $\mathbf{u}(k)$ is the control vector $\mathbf{w}(k)$ represents errors in the command signals and the noise vector $\mathbf{v}(k)$ compensates for modelling errors and errors due to linearization. Both noise components $\mathbf{w}(k)$ and $\mathbf{v}(k)$ are modelled as realizations of independent Gaussian random variables with covariance matrices $\mathbf{\Sigma}_{\mathbf{ww}}(k)$ and $\mathbf{\Sigma}_{\mathbf{vv}}(k)$, respectively. The covariance matrix associated with the state vector is abbreviated by $\mathbf{P}(k)$. We assume that we are given an initial state estimate $\mathbf{z}_0$ with covariance $\mathbf{P}_0$ representing our prior knowledge of the desired parameters.

2.2. Innovation

The starting point of the update step is given by the *a priori* state estimate $\mathbf{z}^{-}(k)$ at time k and $\mathbf{P}^{-}(k)$, the corresponding covariance matrix. These values are either provided by our prior knowledge of the desired parameters ($k = 0$) or by the results of the prediction step in Sec. 2.2 ($k > 0$). Using the current measurement vector $\hat{\mathbf{x}}(k)$, we will compute both the *a posteriori* state estimate $\mathbf{z}^{+}(k)$ and $\mathbf{x}^{+}(k)$, the best estimate of the ideal observation vector $\mathbf{x}(k)$. Least squares optimal estimates will be obtained by minimizing (the time index k is omitted in the following paragraphs to improve readability)

$$\begin{aligned}
&(\mathbf{x}^+ - \hat{\mathbf{x}})^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{ee}}^{-1} (\mathbf{x}^+ - \hat{\mathbf{x}}) + (\mathbf{z}^+ - \mathbf{z}^-)^T (\mathbf{P}^-)^{-1} (\mathbf{z}^+ - \mathbf{z}^-) \\
&= \begin{pmatrix} \mathbf{x}^+ - \hat{\mathbf{x}} \\ \mathbf{z}^+ - \mathbf{z}^- \end{pmatrix}^T \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{ee}} & 0 \\ 0 & \mathbf{P}^- \end{bmatrix}^{-1} \begin{pmatrix} \mathbf{x}^+ - \hat{\mathbf{x}} \\ \mathbf{z}^+ - \mathbf{z}^- \end{pmatrix}
\end{aligned} \tag{4}$$

subject to the nonlinear constraint

$$\mathbf{h}(\mathbf{z}^+, \mathbf{x}^+) = 0. \tag{5}$$

To solve this optimization problem, we linearize Eq. (5) about an operation point $(\check{\mathbf{z}}, \check{\mathbf{x}})$. This operation point should be chosen according to our best knowledge of the true parameters and observations. In a first iteration, we set $(\check{\mathbf{z}}, \check{\mathbf{x}}) = (\mathbf{z}^-, \hat{\mathbf{x}})$ and obtain

$$\mathbf{h}(\mathbf{z}^+, \mathbf{x}^+) \approx \mathbf{A}(\mathbf{z}^+ - \check{\mathbf{z}}) + \mathbf{B}(\mathbf{x}^+ - \check{\mathbf{x}}) + \mathbf{h}(\check{\mathbf{z}}, \check{\mathbf{x}}) \tag{6}$$

$$= \mathbf{A}\mathbf{z}^+ + \mathbf{B}\mathbf{x}^+ + \mathbf{h}(\check{\mathbf{z}}, \check{\mathbf{x}}) - \mathbf{A}\check{\mathbf{z}} - \mathbf{B}\check{\mathbf{x}} \tag{7}$$

$$= \mathbf{A}\mathbf{z}^+ + \mathbf{B}\mathbf{x}^+ + \mathbf{y}, \tag{8}$$

where $\mathbf{A} = -\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{z}}\right|_{\check{z},\check{x}}$, $\mathbf{B} = \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right|_{\check{z},\check{x}}$ and $\mathbf{y} = \mathbf{h}(\check{\mathbf{z}}, \check{\mathbf{x}}) - \mathbf{A}\check{\mathbf{z}} - \mathbf{B}\check{\mathbf{x}}$. Using Lagrangian multipliers, we then need to find the extremum of the cost function

$$J = \begin{pmatrix} \mathbf{x}^+ - \hat{\mathbf{x}} \\ \mathbf{z}^+ - \mathbf{z}^- \end{pmatrix}^T \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{ee}}^{-1} & 0 \\ 0 & (\mathbf{P}^-)^{-1} \end{bmatrix} \begin{pmatrix} \mathbf{x}^+ - \hat{\mathbf{x}} \\ \mathbf{z}^+ - \mathbf{z}^- \end{pmatrix} - 2\boldsymbol{\eta}(\mathbf{A}\mathbf{z}^+ + \mathbf{B}\mathbf{x}^+ + \mathbf{y}). \tag{9}$$

Taking the derivatives of J with respect to the corrected measurements $\mathbf{x}^+$, the corrected state vector $\mathbf{z}^+$, and the Lagrangian multiplier η, we get after some algebra

$$\frac{\partial J}{\partial \mathbf{x}^+} = 0 \quad \Leftrightarrow \quad \mathbf{x}^+ = \hat{\mathbf{x}} + \boldsymbol{\Sigma}_{\mathbf{ee}} \mathbf{B}^T \boldsymbol{\eta} \tag{10}$$

$$\frac{\partial J}{\partial \mathbf{z}^+} = 0 \quad \Leftrightarrow \quad \mathbf{z}^+ = \mathbf{z}^- + \mathbf{P}^- \mathbf{A}^T \boldsymbol{\eta} \tag{11}$$

$$\frac{\partial J}{\partial \boldsymbol{\eta}} = 0 \quad \Leftrightarrow \quad \mathbf{A}\mathbf{z}^+ + \mathbf{B}\mathbf{x}^+ + \mathbf{y} = \mathbf{0}. \tag{12}$$

Substituting Eqs. (10) and (11) in Eq. (12) yields

$$\left(\mathbf{A}\mathbf{P}^{-}\mathbf{A}^{T} + \mathbf{B}\mathbf{\Sigma}_{\mathbf{ee}}\mathbf{B}^{T}\right)\eta = -\mathbf{A}\mathbf{z}^{-} - \mathbf{B}\hat{\mathbf{x}} - \mathbf{y}. \tag{13}$$

For simplicity, we define a transformed observation vector

$$\hat{\mathbf{x}}^{*} = -\mathbf{B}\hat{\mathbf{x}} - \mathbf{y}, \quad \mathbf{R}^{*} = \mathbf{B}\mathbf{\Sigma}_{\mathbf{ee}}\mathbf{B}^{T} \tag{14}$$

and obtain from Eq. (13)

$$\boldsymbol{\eta} = \left(\mathbf{A}\mathbf{P}^{-}\mathbf{A}^{T} + \mathbf{R}^{*}\right)^{-1}\left(\hat{\mathbf{x}}^{*} - \mathbf{A}\mathbf{z}^{-}\right). \tag{15}$$

The inverse matrix in Eq. (15) exists if all matrices $\mathbf{A}$, $\mathbf{B}, \mathbf{\Sigma}_{\mathbf{ee}}$ and $\mathbf{P}^{-}$ have full rank. This is always true in our application. Combining Eqs. (15), (10) and (11), we have derived the formulas for computing our corrected observations and state parameters

$$\mathbf{x}^{+} = \hat{\mathbf{x}} + \mathbf{\Sigma}_{\mathbf{ee}}\mathbf{B}^{T}\left(\mathbf{A}\mathbf{P}^{-}\mathbf{A}^{T} + \mathbf{R}^{*}\right)^{-1}\left(\hat{\mathbf{x}}^{*} - \mathbf{A}\mathbf{z}^{-}\right) \tag{16}$$

$$\mathbf{z}^{+} = \mathbf{z}^{-} + \mathbf{K}\left(\hat{\mathbf{x}}^{*} - \mathbf{A}\mathbf{z}^{-}\right) \tag{17}$$

$$= \mathbf{z}^{-} + \mathbf{K}\mathbf{h}\left(\mathbf{z}^{-}, \hat{\mathbf{x}}\right), \tag{18}$$

where

$$\mathbf{K} = \mathbf{P}^{-}\mathbf{A}^{T}\left(\mathbf{A}\mathbf{P}^{-}\mathbf{A}^{T} + \mathbf{R}^{*}\right)^{-1}. \tag{19}$$

From Eq. (17) and noting that $\mathbf{R}^{*}$ is the covariance matrix of the transformed observation $\hat{\mathbf{x}}^{*}$, we can easily determine the covariance matrix $\mathbf{P}^{+}$ associated with the updated state estimate

$$\mathbf{P}^{+} = (\mathbf{I} - \mathbf{K}\mathbf{A})\,\mathbf{P}^{-}\left(\mathbf{I} - \mathbf{K}\mathbf{A}^{T}\right) + \mathbf{K}\mathbf{R}^{*}\mathbf{K}^{T}. \tag{20}$$

Sparse matrix operations can be used to implement Eqs. (16-20) efficiently.
To find a better state estimate, we can now use $(\mathbf{z}^{+}, \mathbf{x}^{+})$ as a new operation point $(\check{\mathbf{z}}, \check{\mathbf{x}})$ for the linearization of the nonlinear constraint function $\mathbf{h}(.)$. This process may be repeated until the difference between two refined state estimates falls below a predefined threshold or a fixed number of iterations is reached. In [3], experimental evaluations showed that more than 60% of the improvement by iterated innovation steps is already obtained after one additional linearization. Furthermore, estimation results did not change significantly after more than four iterated linearizations. We have thus chosen to use two additional iterations for the examples in this paper. A summary of the iteration formulas is given in Tab. 1.

Another approach to deal with implicit measurement constraints is the normalized Sampson error [11]. The main difference to our approach, however, is the estimation of an ideal measurement vector. As will be shown in Sec. 3.2, the iterated linearization about both corrected state and observation provides results with higher accuracy as compared to a Taylor approximation of $\mathbf{h}(.)$ about an updated state estimate and the noise measurements $\hat{\mathbf{x}}$.

2.3. Prediction

Given the *a posteriori* state estimate $\mathbf{z}^{+}(k)$ and $\mathbf{P}^{+}(k)$ as provided by the innovation step, we can predict the *a priori* state vector $\mathbf{z}^{-}(k+1)$ and its covariance matrix $\mathbf{P}^{-}(k+1)$ in the next time step $k+1$. This is achieved using the standard formulas of an extended Kalman filter (e.g. [4,12]):

$$\mathbf{z}^{-}(k+1) = \mathbf{f}\left(\mathbf{z}^{+}(k), \mathbf{u}(k)\right) \tag{21}$$

$$\mathbf{P}^{-}(k+1) = \mathbf{F}\mathbf{P}^{+}(k)\,\mathbf{F}^{T} + \mathbf{G}\Sigma_{\mathbf{ww}}(k)\,\mathbf{G}^{T} + \Sigma_{\mathbf{vv}}(k). \tag{22}$$

where $\mathbf{F} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{z}}\right|_{\mathbf{z}^{+}(k),\mathbf{u}(k)}$ and $\mathbf{G} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{u}}\right|_{\mathbf{z}^{+}(k),\mathbf{u}(k)}$.

We have now derived an IEKF for observations models with nonlinear constraints that not only computes an improved state vector but also an estimate of the true observation vector for subsequent linearization. Tab.1 summarizes our algorithm.

3. EXTRINSIC AUTO-CALIBRATION OF STEREO CAMERAS

3.1. Stereo self-calibration

In this section, we give a short outline of our algorithm to determine extrinsic camera parameters (i.e. the orientations of both cameras with respect to a fixed base line) automatically. A more detailed description can be found in [3]. The nonlinear constraint equations have also been presented in [2], although without an correction of the measurement vector for iterated linearizations as derived in this article.

Fig. 1 illustrates the camera parameters that have to be estimated by our stereo auto-calibration. The extrinsic camera parameters are given by yaw, pitch, and roll angles $(\Psi_L, \Phi_L, \Theta_L)$ for the left and $(\Psi_R, \Phi_R = 0, \Theta_R)$ for the right camera, respectively. Please note that $\Phi_R = 0$ since the world coordinate system (WCS) was chosen such that the viewing direction of the right camera is in the x-y-plane of the WCS. The base length b cannot be determined from image data alone without a known reference length or known absolute velocity of the observer. In fact, the base length is usually known quite well from the production process of the stereo rig while the orientations are hard to determine before hand and are required with an accuracy of about 10^{-2}

degrees in some applications. Therefore, we do not recover b in this application. In addition, since we will use the 2D motion of image points between two consecutive images of the same camera as an input to our calibration

Table 1. IEKF for observation models with nonlinear constraints.

System model: $\mathbf{z}_k = \mathbf{f}\left(\mathbf{z}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}\right)$, $\quad \mathbf{w}_k \sim N(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{ww}})$.
Observation model: $\hat{\mathbf{x}}_k = \mathbf{x}_k + \mathbf{e}_k$, $\quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{z}_k\right) = \mathbf{0}$, $\quad \mathbf{e}_K \sim N(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{ee}})$.
Assume that the conditions of Sec. 2.1 hold.
Initial parameter vector and its covariance: $\hat{\mathbf{z}}_0^- = \mathbf{z}_0$, $\mathbf{P}_0^- = \mathbf{P}_0$.
Let $k=0$.

1. Update step

$$\check{\mathbf{x}}_{k,0} = \hat{\mathbf{x}}_k \,, \quad \check{\mathbf{z}}_{k,0} = \hat{\mathbf{z}}_k^+ \,.$$

For $l = 0 \ldots L-1$:

$$\mathbf{A}_l = \partial \mathbf{h}/\partial \mathbf{z}\,|_{\check{\mathbf{x}}_{k,l}, \check{\mathbf{z}}_{k,l}} \,, \quad \mathbf{B}_l = \partial \mathbf{h}/\partial \mathbf{x}\,|_{\check{\mathbf{x}}_{k,l}, \check{\mathbf{z}}_{k,l}} \,.$$
$$\mathbf{K}_{k,l} = \mathbf{P}_k^- \mathbf{A}_l^{\mathrm{T}} \left[\mathbf{A}_l \mathbf{P}_k^- \mathbf{A}_l^{\mathrm{T}} + \mathbf{B}_l \mathbf{\Sigma}_{\mathbf{ee}} \mathbf{B}_l^{\mathrm{T}}\right]^{-1} \,.$$
$$\mathbf{r}_{k,l} = \mathbf{h}\left(\check{\mathbf{x}}_{k,l}, \check{\mathbf{z}}_{k,l}\right) + \mathbf{B}_l \cdot \left(\hat{\mathbf{x}} - \check{\mathbf{x}}_{k,l}\right) + \mathbf{A}_l \cdot \left(\hat{\mathbf{z}}_k^- - \check{\mathbf{z}}_{k,l}\right) \,.$$
$$\check{\mathbf{z}}_{k,l+1} = \hat{\mathbf{z}}_k^- - \mathbf{K}_{k,l}\, \mathbf{r}_{k,l} \,.$$
$$\check{\mathbf{x}}_{k,l+1} = \hat{\mathbf{x}}_k - \mathbf{\Sigma}_{\mathbf{ee}} \mathbf{B}_l^{\mathrm{T}} \left[\mathbf{B}_l \mathbf{\Sigma}_{\mathbf{ee}} \mathbf{B}_l^{\mathrm{T}}\right]^{-1} \mathbf{r}_{k,l} \,.$$

$$\hat{\mathbf{z}}_k^+ = \check{\mathbf{z}}_{k,L} \,.$$
$$\mathbf{P}_k^+ = \left[\mathbf{I} - \mathbf{K}_{k,L-1} \mathbf{A}_{L-1}\right] \mathbf{P}_k^- \,.$$

2. Prediction step:

$$\mathbf{F}_{\mathbf{z}} = \partial \mathbf{f}/\partial \mathbf{z}\,|_{\hat{\mathbf{z}}_k^+, \mathbf{u}_k, \mathbf{0}} \,, \quad \mathbf{F}_{\mathbf{w}} = \partial \mathbf{f}/\partial \mathbf{w}\,|_{\hat{\mathbf{z}}_{k1}^+, \mathbf{u}_{k1}, \mathbf{0}} \,.$$
$$\hat{\mathbf{z}}_{k+1}^- = \mathbf{f}\left(\hat{\mathbf{z}}_k^+, \mathbf{u}_k, \mathbf{0}\right) \,.$$
$$\mathbf{P}_{k+1}^- = \mathbf{F}_{\mathbf{z}}\, \mathbf{P}_k^+\, \mathbf{F}_{\mathbf{z}}^{\mathrm{T}} + \mathbf{F}_{\mathbf{w}}\, \mathbf{\Sigma}_{\mathbf{ww}}\, \mathbf{F}_{\mathbf{w}}^{\mathrm{T}} \,.$$

3. Set $k := k+1$. Go to step 1.

algorithm, we also need to compute the 3D motion of the stereo rig. $(\Psi_M, \Phi_M, \Theta_M)$ and $(\upsilon_x, \upsilon_y, \upsilon_z)$ denote the rotational and translational components of the observer's motion, respectively. Our desired parameter is thus given by

$$\mathbf{z} = (\Psi_L, \Phi_L, \Theta_L, \Psi_R, \Theta_R, \Psi_M, \Phi_M, \Theta_M, \upsilon_x, \upsilon_y, \upsilon_z)^{\mathrm{T}}. \tag{23}$$

To formulate an observation model, we use epipolar and trilinear constraints as described in e.g. [2]:

– Epipolar constraint: Let $\mathbf{x}_L = (x_L, y_L, 1)$ and $\mathbf{x}_R = (x_R, y_R, 1)$ denote the homogeneous image coordinates of an object point $\mathbf{X}$ in the left and right stereo image, respectively. These projections are related by the epipolar constraint

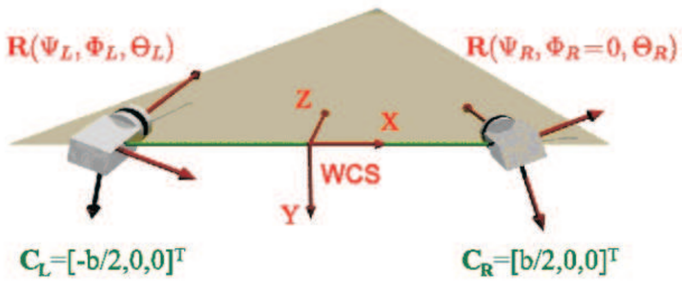

Fig. 1. Camera Model. The world coordinate system (WCS) is located in the middle of the baseline of the stereo cameras. Additionally, we impose that the X-axis of the WCS is aligned with the base line and the optical axes of the right camera is in the X-Y-plane. b denotes the base length of the stereo rig and the rotation matrices are specified in yaw, pitch and roll angles.

$$\mathbf{x}_L^{\mathrm{T}}\mathbf{K}_L^{-T}\mathbf{R}_L\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & -b\\ 0 & b & 0\end{pmatrix}\mathbf{R}_R^{\mathrm{T}}\mathbf{K}_R^{-1}\mathbf{x_R} = 0 \tag{24}$$

where $\mathbf{R}_L = \mathbf{R}(\Psi_L, \Phi_L, \Theta_L)$ and $\mathbf{R}_R = \mathbf{R}(\Psi_L, 0, \Theta_L)$ indicate the 3×3-rotation matrices of both cameras with respect to the base line, and $\mathbf{K}_L$ and $\mathbf{K}_R$ specify the intrinsic matrices of the left and right camera, respectively. Using e.g. correlation based block matching techniques, we can find (noisy) correspondence pairs in each frame of the stereo rig:

$$\begin{pmatrix}\hat{x}_L\\ \hat{y}_L\end{pmatrix} = \begin{pmatrix}x_L\\ y_L\end{pmatrix} + \begin{pmatrix}e_{x,L}\\ e_{y,L}\end{pmatrix}, \quad \begin{pmatrix}\hat{x}_R\\ \hat{y}_R\end{pmatrix} = \begin{pmatrix}x_R\\ y_R\end{pmatrix} + \begin{pmatrix}e_{x,R}\\ e_{y,R}\end{pmatrix} \tag{25}$$

with errors $(e_{x,L}, e_{y,L})^{\mathrm{T}}$ and $(e_{x,R}, e_{y,R})^{\mathrm{T}}$. Eqs. (24) and (25) define an observation model as formulated in Sec. 2.1.

– *Trilinear constraints*: If – in addition to the extracted stereo matches $\mathbf{x_L}, \mathbf{x_R}$ – we can find a corresponding image point $\mathbf{x_{R+}}$ in the next right image of the sequence (i.e. a temporal match), we are able to employ trilinear constraints. These constraints allow stereo-self calibration with higher accuracy than the epipolar constraint. The trilinear constraints have the following form

$$\mathbf{h}(\Psi_L, \Phi_L, \Theta_L, \Psi_R, \Theta_R, \Psi_M, \Phi_M, \Theta_M, \upsilon_x, \upsilon_y, \upsilon_z, \mathrm{x}_L, \mathrm{x}_R, \mathrm{x}_{R+}) = \mathbf{0} \tag{26}$$

with observations

$$\begin{gathered}\begin{pmatrix}\hat{x}_L\\ \hat{y}_L\end{pmatrix} = \begin{pmatrix}x_L\\ y_L\end{pmatrix} + \begin{pmatrix}e_{x,L}\\ e_{y,L}\end{pmatrix}, \quad \begin{pmatrix}\hat{x}_R\\ \hat{y}_R\end{pmatrix} = \begin{pmatrix}x_R\\ y_R\end{pmatrix} + \begin{pmatrix}e_{x,R}\\ e_{y,R}\end{pmatrix},\\ \begin{pmatrix}\hat{x}_{R+}\\ \hat{y}_{R+}\end{pmatrix} = \begin{pmatrix}x_{R+}\\ y_{R+}\end{pmatrix} + \begin{pmatrix}e_{x,R+}\\ e_{y,R+}\end{pmatrix}.\end{gathered} \tag{27}$$

The definition of the exact formulas for the trilinear constraints is beyond the scope of this paper. The reader is referred to[3].

We can combine all available correspondence features by simply stacking the epipolar and trilinear constraint equations as well as the corresponding measurement vectors. The state transition function in this application is given by an identity transformation. If we are given a model of the observer's motion, this information can be used here. We have thus defined the implicit observation and the system model for the extrinsic self-calibration of a stereo camera. Results on simulated data and real imagery will be presented in the next section.

3.2. Experimental results

To evaluate the effect of linearizing the nonlinear observation constraint equation about both the estimated parameter vector as well as the corrected measurements (cf. Sec. 2), two different versions of the self-calibration method were tested on synthetic image data: The first version, "with corrected observations", linearized the constraint function about both the current parameter estimate and the corrected measurement vector $\mathbf{x}^+(k)$. The second version, "without corrected observations", performed linearization about the updated state vector and the uncorrected measurements $\hat{\mathbf{x}}(k)$. To generate artificial stereo and 2D motion correspondence data, we simulated stereo image sequences of a moving 3D point cloud with 40 points. The generated input data was perturbed with additive Gaussian white noise of standard deviation pixels and fed into the auto-calibration algorithm. At each time step, the resulting camera parameters were used to compute the 3D positions $\hat{\mathbf{X}}_i$ of the observed image points using Hartley's triangulation method. This enabled us to compute a relative 3D reconstruction error $\epsilon_i = \|\hat{\mathbf{X}}_i - \mathbf{X}_i\|/\|\mathbf{X}_i\|$ with respect to the true 3D position $\mathbf{X}_i$. Fig. 2 depicts the results of 40 independent simulation runs. We found that "with corrected observations" yields significantly better results than "without corrected observations". The advantage seems to become dominant when the current state estimate is closer to the true state vector. After 40 frames, the difference between the mean relative reconstruction errors is approx. 5% indicating that our algorithm can improve the accuracy of stereo self-calibration.

To show the feasibility of our auto-calibration method on real imagery, we conducted experiments with a hand held stereo-rig in a laboratory environment. Fig. 3 shows typical images of our evaluations. Stereo and temporal matches were extracted using a Harris corner detector [5] and standard block matching (e.g. [1]). Stereo depth images were computed using the recovered camera parameters and the stereo matching procedure as described in [6]. The lower left image of Fig. 3 shows the stereo disparity map at $k = 0$ which corresponds to the initial guess used to start the auto-calibration. Here, a reliable 3D reconstruction was not possible. After 40 frames, our algorithm provided reliable camera parameters which allowed to compute a dense depth map. A textured 3D reconstruction of the observed scene at $k = 40$ is shown in Fig. 4.

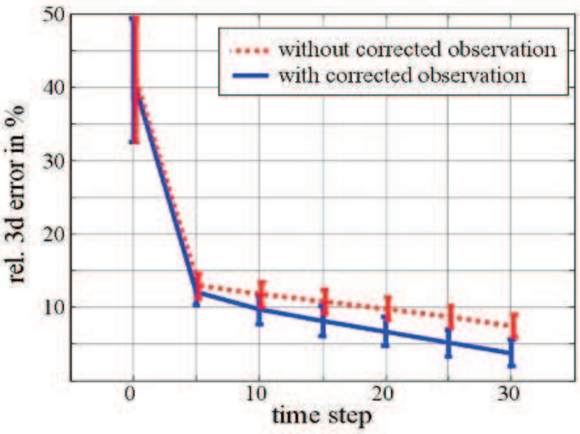

Fig. 2. Recursive self-calibration with and without observation correction. The figure shows the mean and standard deviations of 40 independent simulations runs. The linearization about both the current parameter estimate and the corrected observation vector yields an improvement of approx. 5% in the final relative 3D reconstruction error as compared to a standard IEKF algorithm.

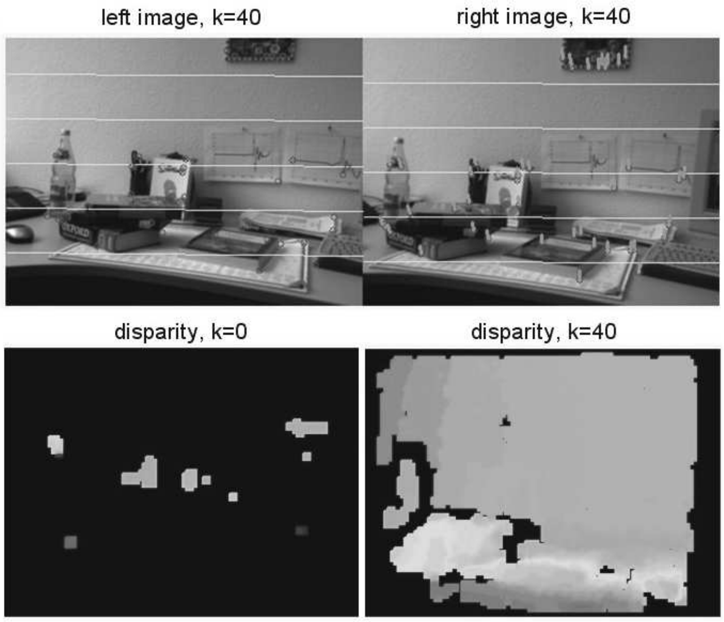

Fig. 3. Self-calibration result after 40 frames. Top: The left and right input images. Extracted features are indicated by red circles (stereo matches) and green lines (temporal matches). Yellow lines show corresponding epipolar lines. Bottom: Color coded disparity map (red color indicates close distance to the observer) at frames $k = 0$ (initial guess) and $k = 40$. Note that the initial guess did not allow a reliable stereo reconstruction while a dense disparity map is obtained using the self-calibration results.

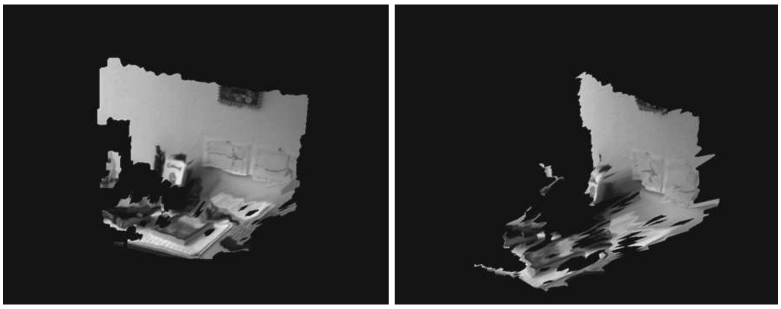

Fig. 4. 3D reconstruction after self-calibration.

4. CONCLUSION

This paper presents a compact derivation of an extended iterated Kalman filter (IEKF) for observation models with nonlinear constraints. It is shown that such a recursive algorithm should not only compute estimates of the desired parameters, but also of the ideal measurements. As compared to standard IEKF approaches, our algorithm can provide higher accuracy that justifies the additional computational load.

The filter algorithm was utilized for the extrinsic auto-calibration of a stereo rig and provided very good results on synthetic and real imagery. However, implicit observation constraints are encountered in a variety of computer vision problems as e.g. structure-from-motion or tracking algorithms. We believe the presented recursive estimation method is applicable to a wide range of applications beyond auto-calibration.

REFERENCES

1. Aschwanden P., Guggenbühl W.: "Experimental results from a comparative study on correlation-type registration algorithms". *In Förstner and Ruwiedel, editors, Robust Computer Vision*, 1992, pp. 268–282.
2. Dang T., Hoffmann C.: "Tracking camera parameters of an active stereo rig". *In 28th Annual Symposium of the German Association for Pattern Recognition (DAGM 2006)*, Berlin, September 12-14 2006.
3. Dang T.: „Kontinuierliche Selbstkalibrierung von Stereokameras". *PhD thesis*, Universität Karlsruhe (TH), 2007.
4. Gelb A.: *Applied Optimal Estimation. MIT Press, Massachusetts Institute of Technology*. Cambridge, MA, 1994.
5. Harris C., Stephens M.J.: "A combined corner and edge detector". *In Alvey 88*, 1988, pp. 147–152.
6. Hirschmüller H., Innocent P.R., Garibaldi J. M.: "Real-time correlation-based stereo vision with reduced border errors". *International Journal of Computer Vision*, 2002, 47(1-3):229–246.
7. Horaud R., Csurka G., Demirdijian D.: "Stereo calibration from rigid motions". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(12):1446–1452.
8. Maybeck P. S.: "Stochastic Models". *Estimation and Control*, vol. 1, 1982.

9. Pollefeys M.: “Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences”. *PhD thesis*, Katholieke Universiteit Leuven, Belgium, 1999.
10. Ressl C.: “Geometry, Constraints and Computation of the Trifocal Tensor”. *PhD thesis*, Technische Universität Wien, Fakultät für Naturwissenschaften und Informatik, 2003.
11. Sampson P.D.: “Fitting conic sections to very scattered data: An iterative refinement of the bookstein algorithm”. *Computer Graphics and Image Processing*, 1982, 18:97–108.
12. Simon D. : *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*. John Wiley & Sons, 2006.
13. Soatto S., Frezza R., Perona P.: “Motion estimation via dynamic vision”. *IEEE ctions on Automatic Control,* 1996, 4(3):393–413,.
14. Tsai R. Y.: “A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses”. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344.
15. Zhang Z., Faugeras O.: “3D Dynamic Scene Analysis, of Springer Series in Information Sciences”. *Springer,* vol. 27, 1992.
16. Zhang Z., Luong Q., Faugeras O.: “Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction:. *IEEE Transactions on Robotics and Automation,* 1996, 12:103–113.
17. Zisserman A., Beardsley P., Reid I.: “Metric calibration of a stereo rig”. *In IEEE Workshop on Representations of Visual Scenes*, 1995, pp. 93–100.