

# Recursive Structure from Motion using Hybrid Matching Constraints with Error Feedback <sup>\*</sup>

Fredrik Nyberg and Anders Heyden

Applied Mathematics Group  
School of Technology and Society  
Malmö University, Sweden  
{fredrik.nyberg, heyden}@ts.mah.se

**Abstract.** We propose an algorithm for recursive estimation of structure and motion in rigid body perspective dynamic systems, based on the novel concept of continuous-differential matching constraints for the estimation of the velocity parameters. The parameter estimation procedure is fused with a continuous-discrete extended Kalman filter for the state estimation. Also, the structure and motion estimation processes are connected by a reprojection error constraint, where feedback of the structure estimates is used to recursively obtain corrections to the motion parameters, leading to more accurate estimates and a more robust performance of the method. The main advantages of the presented algorithm are that after initialization, only three observed object point correspondences between consecutive pairs of views are required for the sequential motion estimation, and that both the parameter update and the correction step are performed using linear constraints only. Simulated experiments are provided to demonstrate the performance of the method.

## 1 Introduction

Structure from motion is one of the central problems in computer vision and has been extensively studied during the last decade. Given a sequence of 2-D images obtained using a single moving camera, the objective is to compute the motion of the camera and a 3-D model of the observed scene. The standard method is to first estimate the motion of the camera, based on matching tensors, obtained from point correspondences in a discrete image sequence. Then, given the motion of the camera, the structure of the scene is obtained as a sparse set of 3-D points, which can be used as a starting point for surface estimation or texture mapping, c.f. [2].

The most common method for estimation of the matching constraints is based on a discrete setting, where e.g. the fundamental (or essential) matrix is estimated between an initial view and another view obtained later in the sequence, c.f. [4]. Another approach, closely related to optical flow, is to use

---

<sup>\*</sup> This work was partially supported by the SRC project 621-2002-4831

a continuous setting and estimate the motion parameters from continuous time matching constraints based on image point positions and velocities, c.f. [5,11,13].

A large class of algorithms utilizes a dynamic systems formulation for the purpose of estimation. The quantities to be estimated are then expressed as states or parameters of a dynamic system, and the estimation task is posed as a problem of state or parameter estimation. The resulting estimation algorithms typically perform recursive estimation, where the estimated variables are updated each time a new image is processed. The use of dynamic systems theory for estimation of motion is described e.g. in [6, 10]. Estimation of structure using dynamic systems is described e.g. in [9], which contains results regarding observability, and also presents algorithms and experimental results.

Attempts has been made to combine the discrete and the continuous methods. In [12], a number of differential matching constraints were derived and an algorithm for updating the fundamental matrix along an image sequence was outlined. However, no experimental evidence or details about the algorithm were given. In this work we derive and utilize a novel matching constraint, called the *continuous-differential epipolar constraint* (CDEC), for the estimation of motion parameters. The CDEC is here fused with a continuous-discrete extended Kalman filter for the state estimation, in order to construct an algorithm for recursive estimation of both structure and motion in a rigid body perspective system. The system is here formulated as a set of ordinary differential equations describing the motion of an object, combined with a measurement equation in the form of perspective observations given at discrete time instants. We also introduce a linear reprojection error constraint, where feedback of the structure estimates is used to recursively obtain corrections to the motion estimates. This constraint connects the structure and motion estimation processes in a consistent way, and is shown by simulations to significantly improve the performance of the method.

The main advantages of the CDEC-based method is that three image correspondences between consecutive image pairs are sufficient for the motion parameter estimation, and that the update, as well as the parameter refinement step, are performed using linear constraints only. This is clearly an advantage e.g in real time applications, where computational speed and memory size might be important performance factors.

## 2 Perspective system

### 2.1 Motion models

We consider two coordinate frames. One coordinate frame, the inertial world coordinate frame, is considered fixed, while the other frame is assumed to be attached to a moving camera, with its origin located at the camera center and its third coordinate axis aligned with the optical axis. Without loss of generality, the moving camera frame can be assumed to coincide with the inertial frame at some initial time instant  $t_0$ .

The relative motion between two camera positions is assumed to be described by a rigid body transformation, obtained using a rotation matrix  $R \in SO(3)$  and a translation vector  $b \in \mathbb{R}^3$ . Then, given a 3-D point on an observed object, its inertial coordinates  $\mathbf{X}_{t_0}$  and  $\mathbf{X}_t$  relative to the camera positions at time  $t_0$  and  $t$  respectively, are related by

$$\mathbf{X}_t = R_t \mathbf{X}_{t_0} + b_t . \quad (1)$$

Assuming  $R$  and  $b$  to be sufficiently smooth as functions of time, the same relation can be described in the continuous case using a dynamic systems formulation as

$$\dot{\mathbf{X}}(t) = \hat{w}(t)\mathbf{X}(t) + \nu(t) , \quad \mathbf{X}(t_0) = \mathbf{X}_{t_0} , \quad (2)$$

where  $\nu \in \mathbb{R}^{3 \times 1}$  is defined by the relation

$$\nu(t) = \dot{b}(t) - \hat{w}(t)b(t) , \quad b(t_0) = b_{t_0} , \quad (3)$$

and  $\hat{w} \in so(3)$  is the skew-symmetric matrix

$$\hat{w} = \dot{R}R^T , \quad (4)$$

parameterized by the vector  $w = (w_1 \ w_2 \ w_3)^T$  according to

$$\hat{w} = \begin{bmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{bmatrix} . \quad (5)$$

Note that, for a constant  $w$ , we get by (4) that  $R_t = e^{\hat{w}t}R_{t_0}$ , where  $R_{t_0}$  is the initial configuration. Also, if we let  $R_{t_1}$  denote the rotation between times  $t_0$  and  $t_1$ , and let  $R_{\Delta t}$  denote the rotation between times  $t_1$  and  $t_2$ , then the rotation  $R_{t_2}$  between times  $t_0$  and  $t_2$  is given by the composition  $R_{t_2} = R_{\Delta t}R_{t_1}$ . Further note that the exponential matrix  $e^{\hat{w}t}$  can be efficiently computed using Rodrigues' formula [8].

In (2),  $w(t)$  can be interpreted as the angular velocity of the camera in the inertial coordinates. The additive component  $\nu(t)$  on the other hand, describes the translational velocity of an imaginary point attached to the moving camera frame, traveling through the origin of the inertial system at time  $t$ , c.f. [8].

Note that since we normally are able to observe only the *relative* motion between the object and the camera, the system (2) can also be interpreted as describing the motion of a point attached to a moving rigid body, as viewed from an inertial coordinate frame located at the center of a fixed camera.

## 2.2 Image acquisition models

Assuming a calibrated standard pinhole camera model [2], homogenous image coordinates  $\mathbf{x}_{t_i}$  of an observed object point  $\mathbf{X} = \mathbf{X}_{t_0}$ , obtained at discrete time instants  $t_i$ , can be described using the relation

$$\lambda_{t_i} \mathbf{x}_{t_i} = R_{t_i} \mathbf{X} + b_{t_i} , \quad i = 0, 1, 2, \dots \quad (6)$$

were  $\lambda_{t_i}$  is a scale factor. We will assume that the object coordinate system has been chosen such that  $R_{t_0} = \mathbf{I}$  and  $b_{t_0} = 0$ .

In the dynamic system formulation corresponding to (2), the images, again assumed to be obtained at discrete time instants  $t_i$  using a calibrated camera, can be described by the relation

$$\lambda_{t_i} \mathbf{x}_{t_i} = \mathbf{X}_{t_i}, \quad i = 0, 1, 2, \dots \quad (7)$$

In the following we make the assumption of normalized homogenous image coordinates. Thus  $\lambda_{t_i} = \mathbf{X}_{3,t_i}$ , i.e. the scale factor  $\lambda_{t_i}$  equals the point depth at time  $t_i$ .

Combining the continuous motion equations (2) with the discrete output equation (7) results in one version of what is sometimes denoted a *dynamic perspective system* [1].

In the following it is assumed without loss of generality that  $t_0 = 0$ , and that the discrete events are equally spaced in time, i.e.  $t_{i+1} - t_i = \Delta t$ , for  $i = 0, 1, 2, \dots$ , and for some small number  $\Delta t > 0$ .

### 2.3 Problem formulation

A structure and motion estimation problem can now be formulated as the task of recursively estimating both the state  $\mathbf{X}(t)$  and the motion parameters  $w(t)$  and  $\nu(t)$  of the system model (2) at the time  $t$ , given the set of perspective measurements  $\mathfrak{M}_t = \{\mathbf{x}_{t_i} \mid \forall i : t_i \leq t\}$ . Or equivalently, given  $\mathfrak{M}_t$ , recover the 3-D position  $\mathbf{X}$  of an observed point on a stationary object and the extrinsic camera parameters  $R_t, b_t$  in the model (1).

## 3 Matching constraints

### 3.1 The standard epipolar constraint

Discrete time matching constraints can be obtained using the relation (6), for several different time instants  $t_i$ , and eliminating the object point coordinates  $\mathbf{X}$  from the resulting system of equations. In this paper we limit ourselves to the two-view constraint, which is thus obtained from

$$\begin{cases} \lambda_0 \mathbf{x}_0 = \mathbf{X} \\ \lambda_t \mathbf{x}_t = R_t \mathbf{X} + b_t, \end{cases} \quad (8)$$

where for simplicity of notation we use  $t = t_i$  for some  $i \in \{1, 2, \dots\}$ . Eliminating  $\mathbf{X}$  from (8) gives the well known *discrete epipolar constraint*

$$\mathbf{x}_0^T R_t^T \widehat{b}_t \mathbf{x}_t = 0, \quad (9)$$

where  $\widehat{b} \in so(3)$  denotes the skew-symmetric matrix corresponding to the vector  $b$  in the same way as  $\widehat{w}$  is related to  $w$  by (5). The matrix  $E_t \doteq R_t^T \widehat{b}_t$  in (9) is usually denoted the *essential matrix*.

If measurements are assumed to be continuously available, continuous time matching constraints can be similarly derived using a continuous version of the camera matrix equation (6) giving the *continuous epipolar constraint*

$$\dot{\mathbf{x}}^T \widehat{\nu} \mathbf{x} + \mathbf{x}^T \widehat{w} \widehat{\nu} \mathbf{x} = 0, \quad (10)$$

with  $\nu$  and  $w$  as in (3) and (4), respectively. For details, see e.g. [6].

### 3.2 The continuous-differential epipolar constraint

We now introduce the *continuous-differential epipolar constraint* (CDEC), which is one type of *hybrid* constraints, i.e. constraints combining continuous and discrete elements, that can be used for matching constraint tracking.

Write down the camera matrix equations (6) for times  $t_0 = 0$ ,  $t_i = t$  and  $t_{i+1} = t + \Delta t$  as follows

$$\left\{ \begin{array}{l} \lambda_0 \mathbf{x}_0 = \mathbf{X} \\ \lambda_t \mathbf{x}_t = R_t \mathbf{X} + b_t \\ \lambda_{t+\Delta t} \mathbf{x}_{t+\Delta t} = R_{t+\Delta t} \mathbf{X} + b_{t+\Delta t} \end{array} \right. \quad (11)$$

To obtain first order approximations to the parameter matrices  $R_{t+\Delta t}$  and  $b_{t+\Delta t}$ , assume  $\widehat{w}$  to be constant  $\widehat{w} \equiv \widehat{w}_t$  over the interval  $[t, t + \Delta t]$ . Using (3) and (4) then results in

$$b_{t+\Delta t} \approx b_t + \dot{b}_t \Delta t = b_t + (\nu_t + \widehat{w}_t b_t) \Delta t. \quad (12)$$

and

$$R_{t+\Delta t} \approx R_t + \dot{R}_t \Delta t = (\mathbf{I} + \widehat{w}_t \Delta t) R_t, \quad (13)$$

Also, as a first order approximation to  $\mathbf{x}_{t+\Delta t}$  take

$$\mathbf{x}_{t+\Delta t} \approx \mathbf{x}_t + \dot{\mathbf{x}}_t \Delta t \approx \mathbf{x}_t + \Delta \mathbf{x}_t, \quad (14)$$

where  $\Delta \mathbf{x}_t$  is the image flow vector  $\Delta \mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-\Delta t}$ . Note that the backward difference, rather than the forward difference, is used here, since we intend to use the resulting constraints for recursive parameter estimation, implying knowledge of image data only up to the current time  $t$ .

Eliminating  $\mathbf{X}$  in (11) using the first equation, expanding until first order in  $\Delta t$  using (13) and (14), and assuming a normalization of the image coordinates such that  $\mathbf{x}_t = (x_t \ y_t \ 1)^T$ , and hence  $\Delta \mathbf{x}_t = (\Delta x_t \ \Delta y_t \ 0)^T$ , results in

$$\underbrace{\begin{bmatrix} R_t \mathbf{x}_0 & \mathbf{x}_t & \mathbf{0}_{3 \times 1} & b_t \\ \widehat{w}_t R_t \mathbf{x}_0 \Delta t & \Delta \mathbf{x}_t & \mathbf{x}_t & (\nu_t + \widehat{w}_t b_t) \Delta t \end{bmatrix}}_{M_{\text{CDEC}}} \begin{bmatrix} -\lambda_0 \\ \lambda_t \\ \lambda_{t+\Delta t} - \lambda_t \\ -1 \end{bmatrix} = \mathbf{0}_{6 \times 1}. \quad (15)$$

The CDEC can thus be compactly expressed by the condition

$$\text{rank} [M_{\text{CDEC}}] < 4. \quad (16)$$

## 4 Recursive structure and motion estimation

### 4.1 Motion estimation using CDEC

To perform recursive estimation of both state and parameters in the system (2) given the perspective output (7), we intend to use the CDEC for sequential velocity parameter update and recursively feed these updates to a separate state estimator.

Expanding the minors of  $M_{\text{CDEC}}$  in (15) and imposing the rank constraint (16), gives the following different constraints in the motion parameters:

- Minors containing the first three rows give the standard epipolar constraint.
- Minors containing two rows out of the first three give linear constraints in  $w_t$  and  $\nu_t$ , in total nine such linear constraints.
- Minors containing the three last rows give nonlinear constraints on the motion parameters.

For our purposes, only constraints of the second type are useful. It turns out that there only exist two linearly independent constraints on the motion parameters from the nine constraints of the second type above. This implies that the estimates of  $w_t$  and  $\nu_t$  can be updated using only three corresponding points from a system of the type

$$M_t \begin{bmatrix} w_t \\ \nu_t \end{bmatrix} = m_t, \quad (17)$$

with  $M_t = M_t(\mathbf{x}_0^k, \mathbf{x}_t^k, \Delta \mathbf{x}_t^k, R_t, b_t) \in \mathbb{R}^{9 \times 6}$  and  $m_t = m_t(\mathbf{x}_0^k, \mathbf{x}_t^k, \Delta \mathbf{x}_t^k, R_t, b_t) \in \mathbb{R}^{9 \times 1}$ , for the point number index  $k \in \{1, 2, 3\}$ . Note that the structure of  $M_t$  and  $m_t$  may easily be set up in advance, and then evaluated for a given set of measurements. There is thus no need to actually compute minors in each step of the algorithm.

Given the new velocity parameter estimates  $w_t$  and  $\nu_t$ , approximate values for  $R_{t+\Delta t}$  and  $b_{t+\Delta t}$  can be computed using  $R_{t+\Delta t} = e^{\hat{w}_t \Delta t} R_t$  and (12) respectively.

This method for motion recovery represents a huge improvement compared to the standard discrete approaches, where five point correspondences give highly nonlinear constraints, and at least eight point correspondences are needed to obtain reasonably simple linear constraints.

Due to the local nature of the approximations employed in the CDEC, the method requires fairly accurate initial values for the parameter estimates. An effective initialization procedure is therefore desirable. In this work we utilize a method based on the continuous epipolar constraint (10), leading to the continuous eight-point algorithm [6]. This means that for the very first step of the estimation process, eight point correspondences are needed between the first two images. But once the initial parameter estimates are obtained, only three point correspondences between consecutive pairs of views are needed for subsequent motion recovery.

## 4.2 State estimation using the continuous-discrete EKF

Given the motion parameters it is possible to employ a number of algorithms for recursive structure recovery. Here we optionally select a continuous-discrete extended Kalman (EKF) filter for the state estimation process [7].

It is suggested in [3] that in using an EKF for state estimation in perspective systems, a coordinate transformation should be performed prior to filtering in order to avoid adverse effects due to the nonlinearity in the measurement signal (7). Defining  $\mathbf{X} = (\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3)^T$ , the transformation used here for this purpose is

$$\mathbf{Z} = \begin{pmatrix} \frac{\mathbf{x}_1}{\mathbf{x}_3} & \frac{\mathbf{x}_2}{\mathbf{x}_3} & \frac{1}{\mathbf{x}_3} \end{pmatrix}^T, \quad (18)$$

which has the effect of linearizing the measurement equation and instead introducing a multiplicative nonlinearity in the state equations. The transformed system formulation employed in the state estimation part of the algorithm is then obtained from (2) and (7), using (18) as

$$\begin{cases} \dot{\mathbf{Z}} = A\mathbf{Z} + \zeta + (\xi^T \mathbf{Z})\mathbf{Z}, & \mathbf{Z}(0) = \mathbf{Z}_0 \\ \mathbf{z}_{t_i} = C\mathbf{Z}_{t_i}, & i = 0, 1, 2, \dots, \end{cases} \quad (19)$$

where  $A \in \mathbb{R}^{3 \times 3}$  and  $\zeta, \xi \in \mathbb{R}^{3 \times 1}$  are new system matrices and vectors, parameterized by the velocity vectors  $w$  and  $\nu$  according to

$$A = \begin{bmatrix} 0 & -w_3 & \nu_1 \\ w_3 & 0 & \nu_2 \\ 0 & 0 & 0 \end{bmatrix}, \quad \zeta = \begin{bmatrix} w_2 \\ -w_1 \\ 0 \end{bmatrix}, \quad \xi = \begin{bmatrix} w_2 \\ -w_1 \\ -\nu_3 \end{bmatrix}, \quad (20)$$

and where we have also introduced the camera-type matrix

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \quad (21)$$

Introducing the estimation error covariance matrix  $P \in \mathbb{R}^{3 \times 3}$ , and the EKF parameter matrices  $R_e \in \mathbb{R}^{2 \times 2}$  and  $Q \in \mathbb{R}^{3 \times 3}$ , the discrete time measurement filter algorithm can now be realized as the solution to two groups of equations. At each  $t_i$ , as a new measurement  $\mathbf{z}_{t_i}$  becomes available, the state estimate  $\tilde{\mathbf{Z}}$  is updated according to

$$K_{t_i} = P_{t_i}^- C^T (C P_{t_i}^- C^T + R_e)^{-1} \quad (22)$$

$$\tilde{\mathbf{Z}}_{t_i}^+ = \tilde{\mathbf{Z}}_{t_i}^- + K_{t_i} (\mathbf{z}_{t_i} - C \tilde{\mathbf{Z}}_{t_i}^-) \quad (23)$$

$$P_{t_i}^+ = P_{t_i}^- - K_{t_i} C P_{t_i}^-, \quad (24)$$

where superscripts  $-$  and  $+$  denote the value of the variable before and after the update respectively. Between measurements the estimates are propagated by integrating the deterministic set of differential equations

$$\dot{\tilde{\mathbf{Z}}} = A\tilde{\mathbf{Z}} + \zeta + (\xi^T \tilde{\mathbf{Z}})\tilde{\mathbf{Z}} \quad (25)$$

$$\dot{P} = FP + PF^T + Q \quad (26)$$

from time  $t_i$  to time  $t_i + \Delta t$  with initial values  $\tilde{\mathbf{Z}}(t_i) = \tilde{\mathbf{Z}}_{t_i}^+$  and  $P(t_i) = P_{t_i}^+$  respectively. Here  $F = F(t, \tilde{\mathbf{Z}})$  denotes the jacobian of (19), evaluated at  $\tilde{\mathbf{Z}}$ , i.e.

$$F = A + (\xi^T \tilde{\mathbf{Z}}) \mathbf{I} + \tilde{\mathbf{Z}} \xi^T . \quad (27)$$

In a stochastic setting, the matrices  $R_e$  and  $Q$  in the equations (22) and (26), represent the covariance matrices of the image and the process noise vectors respectively. Since we here deal with a purely deterministic system,  $R_e$  and  $Q$  can instead be considered as design parameters that can be tuned to improve the estimation process.

### 4.3 Motion estimation refinement by reprojection constraints

For simplicity of notation, in this section we let  $t = t_i$  denote one of the discrete time instants when a new measurement becomes available. Given motion estimates  $R_t$  and  $b_t$  obtained using the CDEC through (17), the measurement  $\mathbf{x}_t$ , and the transformed 3-D estimate  $\tilde{\mathbf{Z}}_t^-$  from the EKF propagation equation (25) between times  $t - \Delta t$  and  $t$ , we seek correction vectors  $\alpha, \beta \in \mathbb{R}^{3 \times 1}$  of small magnitude, such that improved motion estimates  $R_t^+$  and  $b_t^+$  are given by

$$R_t^+ = e^{\hat{\alpha}} R_t , \quad b_t^+ = b_t + \beta . \quad (28)$$

To determine the vectors  $\alpha$  and  $\beta$  we intend to utilize a reprojection constraint. Ideally, this implies that given the *true* initial 3-D position  $\mathbf{X}_0$ , we would like to choose  $\alpha$  and  $\beta$  such that the corrections (28) result in the true image  $\mathbf{x}_t$  as the 3-D point position at time  $t$  is reprojected onto the image plane by (6) according to

$$\lambda_t \mathbf{x}_t = R_t^+ \mathbf{X}_0 + b_t^+ . \quad (29)$$

A workable approximate version of the constraint for use in the current context can be derived as follows. By (1) we have

$$\mathbf{X}_0 = R_t^T (\mathbf{X}_t - b_t) . \quad (30)$$

Combining this relation with (28) and (29), and expanding the rotation matrix  $e^{\hat{\alpha}}$  to the first order in  $\hat{\alpha}$  yields

$$\begin{aligned} \lambda_t \mathbf{x}_t &= e^{\hat{\alpha}} \mathbf{X}_t + (\mathbf{I} - e^{\hat{\alpha}}) b_t + \beta \approx (\mathbf{I} + \hat{\alpha}) \mathbf{X}_t + \hat{\alpha} b_t + \beta \\ &= \mathbf{X}_t + \hat{\alpha} (\mathbf{X}_t + b_t) + \beta = \mathbf{X}_t - \widehat{(\mathbf{X}_t + b_t)} \alpha + \beta . \end{aligned} \quad (31)$$

Under the assumption of homogenous image coordinates, we have that  $\lambda_t = \mathbf{X}_{3,t}$ . Further, since in the estimation process we do not have access to the true value for  $\mathbf{X}_t$ , we instead use the available estimate  $\tilde{\mathbf{X}}_t^-$ , obtained by inverse transformation of the EKF estimate  $\tilde{\mathbf{Z}}_t^-$ , which results in

$$\tilde{\mathbf{X}}_{3,t}^- \mathbf{x}_t = \tilde{\mathbf{X}}_t^- - \widehat{(\tilde{\mathbf{X}}_t^- + b_t)} \alpha + \beta , \quad (32)$$



where everything except  $\alpha$  and  $\beta$  can be considered known. Equation (32) can also be expressed as

$$\begin{bmatrix} \widehat{(\tilde{\mathbf{X}}_t^- + b_t)} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \tilde{\mathbf{X}}_t^- - \tilde{\mathbf{X}}_{3,t}^- \mathbf{x}_t. \quad (33)$$

Dividing through by  $\tilde{\mathbf{X}}_{3,t}^-$ , it is seen that the right hand side of (33) equals the reprojection error  $\tilde{\mathbf{x}}_t - \mathbf{x}_t$ . Thus (33) constitutes a reprojection constraint on  $\alpha$  and  $\beta$ , and hence on the motion through (28).

It can be shown that for at least three different observed object point estimates, the resulting systems corresponding to (33) will provide six independent linear constraints on the six unknown parameters in  $\alpha$  and  $\beta$ . Thus the desired correction terms can be determined from the linear system

$$\underbrace{\begin{bmatrix} \widehat{(\tilde{\mathbf{X}}_t^{1-} + b_t)} & -\mathbf{I} \\ \widehat{(\tilde{\mathbf{X}}_t^{2-} + b_t)} & -\mathbf{I} \\ \vdots \\ \widehat{(\tilde{\mathbf{X}}_t^{N-} + b_t)} & -\mathbf{I} \end{bmatrix}}_{M_{repr}} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{X}}_t^{1-} - \tilde{\mathbf{X}}_{3,t}^{1-} \mathbf{x}_t^1 \\ \tilde{\mathbf{X}}_t^{2-} - \tilde{\mathbf{X}}_{3,t}^{2-} \mathbf{x}_t^2 \\ \vdots \\ \tilde{\mathbf{X}}_t^{N-} - \tilde{\mathbf{X}}_{3,t}^{N-} \mathbf{x}_t^N \end{bmatrix}, \quad N \geq 3. \quad (34)$$

It should be noted that (34) is consistent in the sense that if  $\tilde{\mathbf{X}}_t^{k-} = \mathbf{X}_t^k$  for  $k \in \{1, 2, \dots, N\}$ , the right hand side will become the zero vector and, since  $M_{repr}$  has full rank, then also  $\alpha = \beta = \mathbf{0}$ . Hence, according to (28), no correction of the motion estimates will be made.

As will be exemplified in the simulation section, the inclusion of the reprojection constraint correction step significantly enhances the performance of the estimation procedure, leading to more accurate and robust estimates of both structure and motion.

#### 4.4 Structure and motion algorithm

Using the results of the previous sections, the following algorithm can now be employed for recursive structure and motion recovery:

##### 1. Preparations

- Assume that images are obtained sequentially at time instants  $t_i$ ,  $i = 0, 1, 2, \dots$ , equally spaced by  $\Delta t$ . Also assume some initial values  $\tilde{\mathbf{Z}}_{t_1}^-$  and  $P_{t_1}^-$  for the state vector and the error covariance matrix respectively, as well as appropriate values for the other EKF tuning parameters.
- Given the images at times  $t_0 = 0$  and  $t_1 = \Delta t$  with at least eight point correspondences, get initial parameter estimates  $w_0$  and  $\nu_0$  using e.g. the continuous eight-point algorithm.
- Compute  $R_{t_1} = e^{\tilde{w}_0 \Delta t}$  and  $b_{t_1} = \nu_0 \Delta t$ .

## 2. Estimation loop - for $i = 1, 2 \dots$ do

- Using at least three point correspondences, set up the matching constraint matrices  $M_t$  and  $m_t$  of (17) at time  $t_i$  from image points  $\mathbf{x}_0, \mathbf{x}_{t_i}$ , flow vectors  $\Delta \mathbf{x}_{t_i}$  and the parameter matrices  $R_{t_i}$  and  $b_{t_i}$ .
- Solve the resulting linear system (17) for the new parameter estimates  $w_{t_i}$  and  $\nu_{t_i}$  and update the rotation matrix and the translation vector according to  $R_{t_i+\Delta t} = e^{\hat{w}_{t_i} \Delta t} R_{t_i}$  and  $b_{t_i+\Delta t} = b_{t_i} + (\nu_{t_i} + \hat{w}_{t_i} b_{t_i}) \Delta t$  respectively.
- Apply the EKF update step at time  $t_i$  to get the transformed state estimate  $\tilde{\mathbf{Z}}_{t_i}^+$ , and the corresponding estimation error covariance matrix  $P_{t_i}^+$  from the equations (22)-(24)
- Using  $\tilde{\mathbf{Z}}_{t_i}^+$  and  $P_{t_i}^+$  as local initial values, propagate  $\tilde{\mathbf{Z}}(t)$  and  $P(t)$  according to (25) and (26) over the time interval  $[t_i, t_i + \Delta t]$ , with the parameters  $w_{t_i}$  and  $\nu_{t_i}$
- Recover the 3-D position estimate  $\tilde{\mathbf{X}}$  from  $\tilde{\mathbf{Z}}$  on the time interval  $[t_i, t_i + \Delta t]$  by inverse transformation.
- Find correction vectors  $\alpha$  and  $\beta$  by the reprojection constraint (34), and compute refined motion estimates according to (28).

Note that since we are estimating both structure *and* motion, the estimates are inherently subjected to a scale ambiguity. In the above algorithm the scale issue is resolved by assuming the translational velocity vector  $\nu$  to be of unit length in the initialization procedure. This together with the assumption of normalized image coordinates fixes the scale for the subsequent parameter estimates through (17).

## 5 Simulations

Since the initial parameter values obtained by the initialization process generally can be assumed quite accurate, the truly interesting case will be when one or both of the parameter vectors  $w$  and  $\nu$  are time varying. The CDEC-based method can then be evaluated by its ability to follow the time-variations in the parameters, as well as by its ability to correctly recover the 3-D structure.

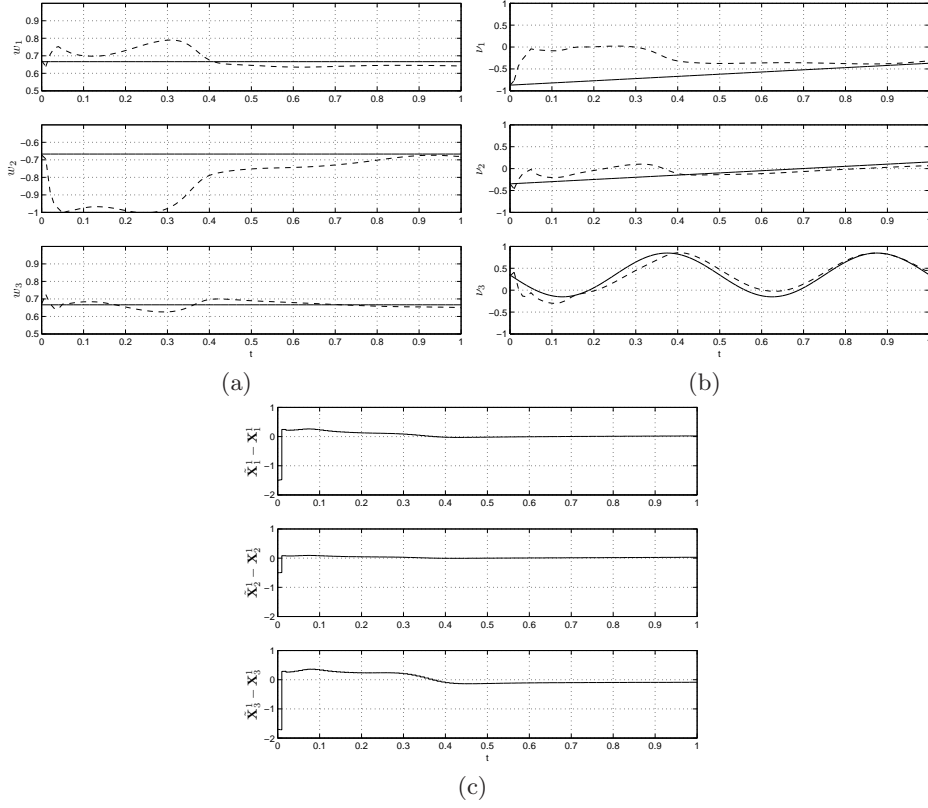
For purpose of illustration we simulate the system (2) for eight points in a general configuration on a grid of stepsize  $\delta t = 10^{-4}$ , and with the parameter vectors

$$w(t) = \frac{2}{3} (1, -1, 1) \quad , \quad \nu(t) = \frac{-1}{\sqrt{1.32}} (1, 0.4, -0.4)^T - \frac{1}{2} (-t, t, \sin 4\pi t)^T \quad .$$

Perspective measurements were computed according to (7) at time instants separated by  $\Delta t = 0.01$ . The true initial values for the three points used throughout the estimation process were

$$\mathbf{X}_0^1 = (1.5, 0.5, 1.7)^T \quad , \quad \mathbf{X}_0^2 = (1, 1.9, 2)^T \quad , \quad \mathbf{X}_0^3 = (1, 1.2, 2.5)^T \quad .$$

The other points were used only in the eight-point algorithm employed in the initialization step. The estimation process was conducted as outlined in Sect.



**Fig. 1.** Estimation results: (a) True (solid) and estimated (dashed) translational velocity  $\nu$ , (b) True (solid) and estimated (dashed) rotational velocity  $w$ , (c) 3-D estimation errors for one of the observed object points.

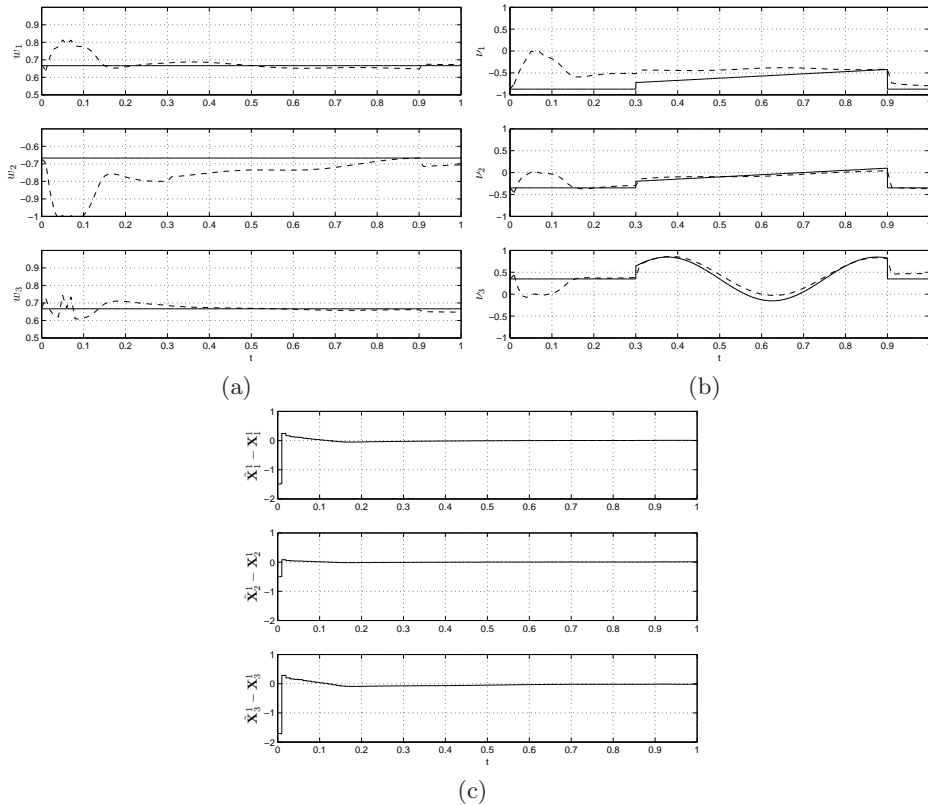
4.4, with the initial transformed state estimates

$$\tilde{\mathbf{Z}}_0^1 = (0.5, 0.5, 0.5)^T, \quad \tilde{\mathbf{Z}}_0^2 = (1, 1, 0.5)^T, \quad \tilde{\mathbf{Z}}_0^3 = (1, 1, 1)^T.$$

The EKF parameters were set to  $P_0 = 100 \cdot \mathbf{I}$ ,  $Q = 0.1 \cdot \mathbf{I}$  and  $R_e = 0.01 \cdot \mathbf{I}$ , for unity matrices  $\mathbf{I}$  of suitable dimensions.

The estimates of the components of the rotational velocity  $w$  and the translational velocity  $\nu$  together with the true values, are shown in Fig. 1(a) and Fig. 1(b) respectively. The resulting 3-D estimation error for one of the observed object points is shown in Fig. 1(c).

To illustrate the fact that the method is able to handle even abrupt changes in the motion parameters, the same system was again simulated but now with the time varying term in the expression for  $\nu$  not turned on until  $t = 0.3$  and again turned off at  $t = 0.9$ . The estimation process was conducted using the same parameters as before, and the results are shown in Fig. 2.

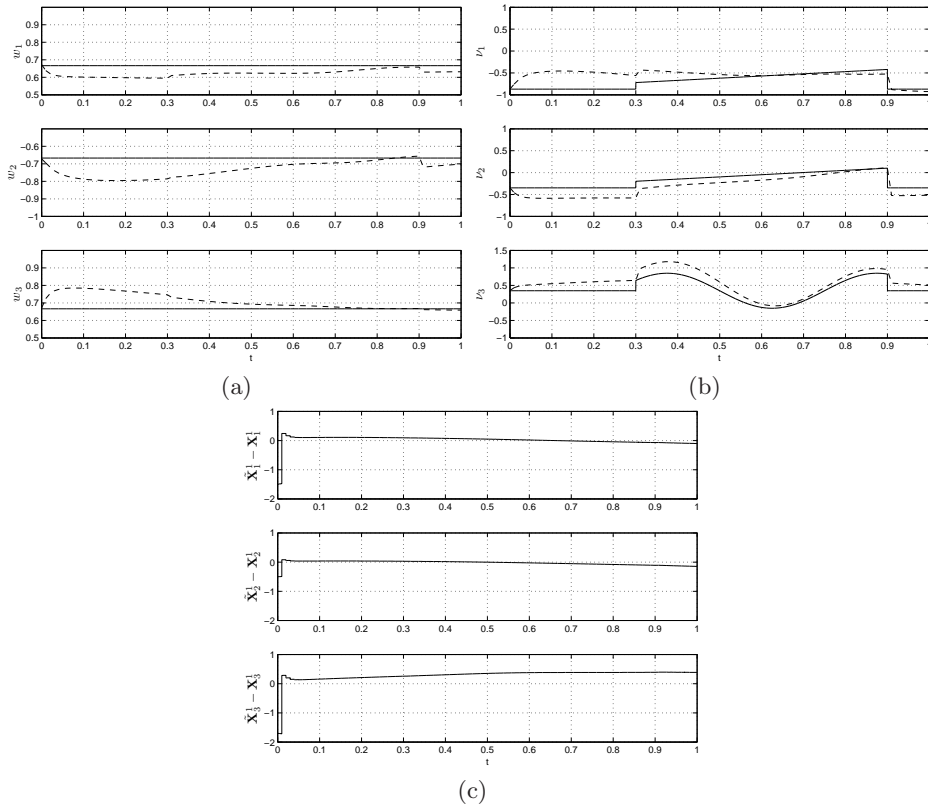


**Fig. 2.** Estimation results for discontinuous motion parameter system: (a) True (solid) and estimated (dashed) translational velocity  $\nu$ , (b) True (solid) and estimated (dashed) rotational velocity  $w$ , (c) 3-D estimation errors for one of the observed object points.

To illustrate the effect of the motion estimate correction step described in Sect. 4.3, the previous experiment was repeated, now *without* using the reprojection constraint. The results are shown in Fig. 3. It can be seen that although the initial estimate in most cases is smoother, the motion estimation in general is now less accurate, and the depth estimation exhibits a distinct bias.

## 6 Conclusion

We have proposed an algorithm for recursive estimation of structure and motion from perspective measurements in a continuous-discrete setting, utilizing the novel concept of the continuous-differential epipolar constraint for the estimation of the velocity parameters, combined with a state estimator, here optionally selected as the continuous-discrete EKF. The structure and motion estimation



**Fig. 3.** Estimation results for discontinuous motion parameter system, obtained without using the correction step: (a) True (solid) and estimated (dashed) translational velocity  $\nu$ , (b) True (solid) and estimated (dashed) rotational velocity  $w$ , (c) 3-D estimation errors for one of the observed object points.

processes are connected by recursive feedback of the structure estimates, resulting in reprojection error constraints used to obtain refined motion estimates. Simulated experiments are included to illustrate the applicability of the concept.

An advantage of the presented method compared to the current state of the art is that once the algorithm has been initialized, using e.g. the continuous eight point algorithm, observations of only three object points are needed for the sequential update and correction of the velocity parameter estimates. Further, after initialization, both the parameter update and the correction step are performed using linear constraints only. A possible drawback of the algorithm is the use of image flow data, which might cause problems when dealing with noisy images.

Note that it is *not* necessary that the same three points are tracked throughout the whole image sequence. By the rigid body assumption, the motion pa-

rameters are common to all points on the observed object. Hence, the only requirement is that three point correspondences are available long enough for the EKF structure estimator to get past the initial transient phase. Then one or all of the point correspondences could be replaced by new ones if necessary, and the estimation process continued, with the appropriate changes made to (15) and consequently to (17).

Future work includes the investigation of the possibility to construct provably convergent adaptive structure and motion estimators based on the continuous-differential matching constraints. Other desirable modifications are extensions of the algorithm to the uncalibrated case, successful filtering of noisy image data and the ability to automatically handle outliers and occlusions. As possible solutions to these issues develop, experiments on real data sequences will be a natural part of future research.

## References

1. Bijoy K. Ghosh, Mrdjan Jankovic, and Y.T. Wu. Perspective problems in system theory and its application to machine vision. *Journal of Mathematical Systems, Estimation and Control*, 4(1):2–38, 1994.
2. Richard Hartley and Andrew Zisserman. *Multiple View Geometry*. Cambridge, 2003.
3. Andreas Huster. *Relative Position Sensing by Fusing Monocular Vision and Inertial Rate Sensors*. PhD dissertation, Stanford University, Department of Electrical Engineering, 2003.
4. H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.
5. Yi Ma, Jana Košecká, and S. Shankar Sastry. Linear differential algorithm for motion recovery. *International Journal of Computer Vision*, 36(1):71–89, 2000.
6. Yi Ma, Stefano Soatto, Jana Košecká, and S. Shankar Sastry. *An Invitation to 3-D Vision*. Springer-Verlag, 2004.
7. P. S. Maybeck. *Stochastic Models, Estimation, and Control Volume 2*. Academic Press, 1982.
8. Richard M. Murray, Zexiang Li, and S. Shankar Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.
9. Stefano Soatto. 3-d structure from visual motion: Modeling, representation and observability. *Automatica*, 33(7):1287–1312, 1997.
10. Stefano Soatto, Ruggero Frezza, and Pietro Perona. Motion estimation via dynamic vision. *IEEE Transactions on Automatic Control*, 41(3), March 1996.
11. Kalle Åström and Anders Heyden. Continuous time matching constraints for image streams. *International Journal of Computer Vision*, 28(1):85–96, 1998.
12. Bill Triggs. Differential matching constraints. In *Proc. of the International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 370–376, 1999.
13. R. Viéville and O. Faugeras. The first order expansion of motion equations in the uncalibrated case. *Computer Vision and Image Understanding*, 64(1):128–146, 1996.