Structure and motion in 3D and 2D from hybrid matching constraints

Anders Heyden, Fredrik Nyberg and Ola Dahl

Applied Mathematics Group
Malmo University, Sweden
{heyden,fredrik.nyberg,ola.dahl}@ts.mah.se

Abstract. *Motion estimation has traditionally been approached either from a pure discrete point of view, using multi-view tensors, or from a pure continuous point of view, using optical-flow-based techniques. This paper builds upon the recent developments of hybrid matching constraints for motion estimation. These hybrid matching constraints are based on both corresponding feature points and the motion of these feature points, thus combining the advantages of both discrete and continuous methods. The main usage of these constraints is as a theoretical basis for filtering approaches to structure and motion recovery, enabling an update of a current motion estimate when a new image becomes available. One important feature is that the update formulas become linear in the motion parameters in the calibrated case, which is a major improvement compared to the standard discrete approach. Another advantage is that fewer points are needed in the update formula than in the traditional discrete case.*
We will present several hybrid matching constraints and derive their properties as well as show how they can be used for structure and motion estimation. First the hybrid bifocal and trifocal constraints will be treated, extending the traditional discrete epipolar and trifocal constraints. Then we will derive novel hybrid constraints for structure and motion recovery from a rigidly moving calibrated stereo-head. Finally, we will derive novel hybrid matching constraints for the 2D-case, enabling linear update of the motion parameters from a calibrated 2D-camera.

1 Introduction

Structure from motion is one of the central problems in computer vision and has been extensively studied during the last decades. The objective is to compute the motion of the camera and the structure of the scene from a number of its two-dimensional images. The standard method is to first estimate the motion of the camera, based on matching tensors, obtained from corresponding points in an image sequence. Then, given the motion of the camera, the structure of the scene is obtained as a sparse set of 3D-points, which can be used as a starting point for surface estimation or texture mapping, cf. [1].

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, cf. [2]. In order to deal with long image sequences several matching constraints are then pasted together, giving a consistent set of matching constraints from which the motion of the camera can be estimated, cf. [3]. Another approach, closely related to optical flow, is to use a continuous setting and estimate the motion parameters from continuous time matching constraints based on image point positions and velocities, cf. [4–6].

Attempts has been made to combine the discrete and the continuous methods. In [7], a number of differential matching constraints were derived and an algorithm for updating the fundamental matrix along an image sequence was indicated. However, no experimental evidence or details about the algorithm were given. Recently, these ideas have been taken up and specified for on-line structure and motion estimation in [8], and preliminary results on an on-line structure and motion system have been reported in [9]. However, these papers are limited to two types of differential matching constraints and the theory is not fully developed.

The main purpose of the present paper is to develop methodology for *on-line recursive structure and motion estimation* for long image sequences. By this, we mean methods that can update a current estimate of the position and orientation of the camera and the structure of the scene, when a new image in the sequence becomes available. Such methods have been presented in [10] and [11], where in both cases complex non-linear procedures are used to update the structure. We will propose a novel method where the motion estimation is separated from the structure estimation, enabling simpler and more stable update schemes.

In this work we derive several types of matching constraint, called *hybrid matching constraints* (HMC), for the estimation and update of the motion parameters. The first one is an extension of the epipolar constraint to a *hybrid epipolar constraint*, where both corresponding points in two images as well as their motion in the second image are used. The second one is an extension of the trifocal constraint to a *hybrid trifocal constraint*, where both corresponding points in three images as well as their motion in the third image are used. The third one is the *hybrid stereo constraint*, where corresponding points together with their motion in a rigidly moving stereo rig are used. Finally, the fourth and fifth ones are the *2D hybrid epipolar constraints* and the *2D hybrid trifocal constraints* where both corresponding points in two and three images respectively as well as their motion in the second and third image respectively are used. All these hybrid constraints will enable us to update the current motion estimate *linearly* based on at least *three corresponding points*. This will be shown theoretically, by proving the exact number of linearly independent constraints obtained from each corresponding point.

The 2D HMC can be used both for pure 2D-cameras, i.e. cameras mapping feature points in a 2D-plane to angles, [12], or to specific cases of 3D perspective structure and motion, e.g. lines in affine cameras and planar motion, [13], that can be reduced to a 2D perspective structure and motion problem.

In order to use these hybrid matching constraints in an on-line structure and motion system, they have to be combined with a structure estimate, e.g. a continuous-discrete extended Kalman filter. Furthermore, a feedback is needed from this structure estimate to refine the motion estimate in the form of a linear reprojection error constraint, cf. [9].

2 Problem Description

2.1 Camera model and notation

We assume the standard pinhole camera model,

$$\lambda \mathbf{x} = P\mathbf{X} \ , \tag{1}$$

where $\mathbf{x}$ denotes homogeneous image coordinates, P the camera matrix, $\mathbf{X}$ homogeneous object coordinates and λ a scale factor. The camera matrix P is usually written as $P = K[\,R \mid -b\,]$, where K denotes the intrinsic parameters and (R, b) the extrinsic parameters (R being an orthogonal matrix). We will from now on assume that the camera is calibrated, i.e. K is known and that the image coordinates have been transformed such that P can be written as $P = [\,R \mid -b\,]$. When several images of the same point are available, (1) can be written as

$$\begin{cases} \lambda(t)\mathbf{x}(t) = P(t)\mathbf{X},\ t \in [0, T] \quad \text{or} \\ \lambda_i \mathbf{x}_i = P_i \mathbf{X},\ i = 1, \dots M \end{cases} \tag{2}$$

in the continuous time case and the discrete time case respectively. The camera matrix P is assumed to have the form

$$P(t) = [\,R(t) \mid -b(t)\,] \quad \text{or} \quad P_i = [\,R_i \mid -b_i\,] \ , \tag{3}$$

in the continuous case and in the discrete case respectively. We furthermore assume that the object coordinate system has been chosen such that $R(0) = R_1 = I$ and $b(0) = b_1 = 0$, implying that $P(0) = P_1 = [\,I \mid 0\,]$.

2.2 Problem formulation

A structure and motion estimation problem can now be formulated as the task of estimating both the structure $\mathbf{X}$ in (2) and the motion parameters $R(t)$ and $b(t)$ in (3) at the time t, given the set of perspective measurements $\mathfrak{M}_t = \{\mathbf{x}(t_i) \mid \forall i \ : \ t_i \leq t\}$. A *recursive* structure and motion problem can be formulated as given an estimate of the structure and motion parameters up to time t, i.e. R_t, b_t and X_t, update this estimate based on measurements obtained up to time $t + \varDelta t$, thus obtaining $R_{t+\varDelta t}$, $b_{t+\varDelta t}$ and $X_{t+\varDelta t}$.

2.3 Discrete Matching Constraints

The *discrete matching constraints* are obtained by using the discrete version of (2), for several different i and eliminating the object coordinates $\mathbf{X}$ and the scale factors λ_i from the resulting equations. In the case of two views we obtain the well-known *epipolar constraint*

$$\mathbf{x}_1^T E \mathbf{x}_2 = 0, \quad \text{with} \quad E = R^T \hat{b} \ , \tag{4}$$

where we for simplicity have used the notation $R_2 = R$ and $b_2 = b$ and $\hat{b}$ denotes the skew-symmetric matrix corresponding to the vector b. The matrix E in (4) denotes the well-known *essential matrix*. This constraint can be used to estimate the motion parameters linearly from at least eight corresponding points. The three- and four-view constraints are obtained similarly, by starting with three (or four) images of the same point and eliminating the object coordinates $\mathbf{X}$, from the resulting system of equations, see [14–16]. In the three-view case the trifocal constraints and the trifocal tensor are obtained and in the four-view case, the quadrifocal constraints and the quadrifocal tensor are obtained. These can be used to estimate the camera motion linearly, from at least 7 and 6 corresponding points respectively, although hidden non-linear constraints are ignored.

2.4 Continuous Matching Constraints

The *continuous time matching constraints* are obtained from the camera matrix equation (2) in continuous form and its time derivative (where for simplicity the time dependency is expressed using an index):

$$
\begin{aligned}
\lambda_t \mathbf{x}_t &= [\, R_t \mid -b_t \,]\mathbf{X} = R_t \tilde{\mathbf{X}} - b_t, \\
\lambda_t' \mathbf{x}_t + \lambda_t \mathbf{x}_t' &= [\, R_t' \mid -b_t' \,]\mathbf{X} = \widehat{w}_t R_t \tilde{\mathbf{X}} - b_t' = \\
&= \hat{w}_t(\lambda_t \mathbf{x}_t + b_t) - b_t' \ ,
\end{aligned}
\tag{5}
$$

where $\tilde{\mathbf{X}}$ denotes the first 3 components of the vector $\mathbf{X}$ and we have assumed that the $\mathbf{X}$ is normalized such that the fourth component is equal to 1. Furthermore, we have used the fact that the derivative of a rotation matrix can be written as $R_t' = \hat{w}_t R_t$, where w_t represents the momentary rotational velocity of the camera at time t. Similarly b_t' denotes the momentary translational velocity of the camera at time t. Define

$$
\nu_t = b_t' - \widehat{w}_t b_t \tag{6}
$$

(representing the momentary translational velocity in a local coordinate system) and multiply the last equation in (5) with $\nu_t \times \mathbf{x}_t$ giving

$$
\mathbf{x}'^T \hat{\nu} \mathbf{x} - \mathbf{x}^T \widehat{w} \hat{\nu} \mathbf{x} = 0 \ , \tag{7}
$$

which is the well known *continuous epipolar constraint.* This constraint can be used to estimate the motion parameters from at least eight corresponding points. Higher order continuous multi-view constraints are obtained by taking higher order derivatives of (2) and then eliminating the structure, cf. [4].

3 Hybrid Matching Constraints

In this section we will derive several *hybrid matching constraints*, that can be used to update the motion parameters recursively. Recall that we assume that estimates of R_t, b_t and X_t are available and that we are looking for methods to update these estimate to time $t + \Delta t$.

3.1 Case I: Epipolar hybrid matching constraints

Write down the camera matrix equations for time 0, time t and time $t + \Delta t$:

$$\begin{cases} \lambda_0 \mathbf{x}_0 = [\, I \mid 0 \,]\mathbf{X} \\ \lambda_t \mathbf{x}_t = [\, R_t \mid -b_t \,]\mathbf{X} \\ \lambda_{t+\Delta t} \mathbf{x}_{t+\Delta t} = [\, R_{t+\Delta t} \mid -b_{t+\Delta t} \,]\mathbf{X} \ . \end{cases} \tag{8}$$

Using $R_t = e^{t\widehat{w}_t}$ as a first order approximation of R, valid between t and $t + \Delta t$, implying that

$$R_{t+\Delta t} = e^{\widehat{w}_t \Delta t} R_t \approx R_t + \widehat{w}_t R_t \Delta t \ ,$$

and

$$\begin{aligned} b_{t+\Delta t} &= b_t + d_t \Delta t, \ \text{corresponding to } d_t = b_t' \ , \\ \lambda_{t+\Delta t} &= \lambda_t + \mu_t \Delta t, \ \text{corresponding to } \mu_t = \lambda_t' \ , \\ \mathbf{x}_{t+\Delta t} &= \mathbf{x}_t + \mathbf{u}_t \Delta t, \ \text{corresponding to } \mathbf{u}_t = \mathbf{x}_t' \ , \end{aligned} \tag{9}$$

and eliminating $\mathbf{X}$ using the first equation in (8) and expanding until the first order in Δt gives

$$\underbrace{\begin{bmatrix} R_t \mathbf{x}_0 & \mathbf{x}_t & 0 & b_t \\ \widehat{w}_t R_t \mathbf{x}_0 & \mathbf{u}_t & \mathbf{x}_t & d_t \end{bmatrix}}_{M_d} \begin{bmatrix} -\lambda_0 \\ \lambda_t \\ \mu_t \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \ , \tag{10}$$

implying that $\operatorname{rank} M_d < 4$, which will be called *the hybrid epipolar constraints*. Assuming normalization such that $\mathbf{x} = (x, y, 1)$, $\mathbf{u} = (u_x, u_y, 0)$, expanding the minors of M_d turns out to give the following different types of constraints in the motion parameters w_t and b_t:

1. Minors containing the first three rows give the discrete epipolar constraint.
2. Minors containing two rows out of the first three give linear constraints in d_t and w_t, in total nine such linear constraints.[1]
3. Minors containing the three last rows give non-linear constraints on the motion parameters.

For our purposes, only the second type of constraints are useful. The first type giving the epipolar constraint can not be used, since we already have an estimate of the motion up to time t and the third ones are not useful because of the non-linearities. In fact there are exactly two linearly independent constraints on the motion parameters from the nine constraints of the second type above, see Appendix A for a detailed proof. This implies that the essential matrix can be updated from at least three corresponding points, which is a huge improvement compared to the standard discrete approaches, where five corresponding points give highly non-linear constraints, and at least eight corresponding points are needed to obtain reasonable simple linear constraints.

[1] There are three ways of omitting one row from the first three rows and then three ways of omitting one row from the last three rows independently.

Observe that when d_t and w_t has been recovered, the new motion parameters and the new essential matrix can easily be obtained from

$$\begin{cases} R_{t+\Delta t} = e^{\widehat{w}_t \Delta t} R_t \\ b_{t+\Delta t} = b_t + d_t \Delta t \end{cases} \quad \Rightarrow \quad E_{t+\Delta t} = R^T_{t+\Delta t} \hat{b}_{t+\Delta t} \ . \tag{11}$$

The update in (11) guarantees that the new essential matrix fulfils the nonlinear constraints. Observe also that although the update is nonlinear in the motion parameters it can be performed efficiently using e.g. Rodriguez formula. Finally, observe that either ν_t or d_t may be used as a parameter for the translational velocity, since they are related according to (6).

3.2 Case II: Trifocal hybrid matching constraints

Write down the camera matrix equations for time 0, s, t and $t + \Delta t$:

$$\begin{cases} \lambda_0 \mathbf{x}_0 = [\, I \mid 0 \,] \mathbf{X} \\ \lambda_s \mathbf{x}_s = [\, R_s \mid -b_s \,] \mathbf{X} \\ \lambda_t \mathbf{x}_t = [\, R_t \mid -b_t \,] \mathbf{X} \\ \lambda_{t+\Delta t} \mathbf{x}_{t+\Delta t} = [\, R_{t+\Delta t} \mid -b_{t+\Delta t} \,] \mathbf{X} \ . \end{cases} \tag{12}$$

Eliminating $\mathbf{X}$ using the first equation and expanding until the first order in Δt give

$$\underbrace{\begin{bmatrix} R_s \mathbf{x}_0 & \mathbf{x}_s & 0 & 0 & b_s \\ R_t \mathbf{x}_0 & 0 & \mathbf{x}_t & 0 & b_t \\ \widehat{w}_t R_t \mathbf{x}_0 & 0 & \mathbf{u}_t & \mathbf{x}_t & d_t \end{bmatrix}}_{N_d} \begin{bmatrix} -\lambda_0 \\ \lambda_s \\ \lambda_t \\ \mu_t \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \ , \tag{13}$$

implying that $\operatorname{rank} N_d < 5$, which will be called *the hybrid trifocal constraints*. The minors of N_d gives the following different constraints in the motion parameters w_t and b_t:

1. Minors containing only one row out of the first three give the previously derived hybrid epipolar constraint.
2. Minors containing only one row out of the last three give the standard discrete trifocal constraints.
3. Minors containing two rows out of the first three, one row out of the middle three rows and two rows out of the last three rows give linear constraints in d_t and w_t, in total 27 such linear constraints.

For our purposes, only the last type of constraints 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, which can be proven in the same way as for the epipolar case in Appendix A. This implies that the trifocal tensor can be updated from at least three corresponding points.

3.3 Case III: Moving stereo head

For simplicity we assume that we have a fixed rigid stereo head, with camera matrices

$$\begin{cases} P_1 = [\, I \mid 0] \,, \\ P_2 = [\, S \mid u] \end{cases} \tag{14}$$

and a rigidly moving scene with coordinates

$$\mathbf{X}(t) = R_t \mathbf{X}_0 + b_t \quad \text{implying} \quad \mathbf{X}'(t) = \widehat{w}_t R_t \mathbf{X}_0 + d_t \tag{15}$$

Write down the camera matrix equations and their derivatives for time t:

$$\begin{cases} \lambda_1 \mathbf{x}_1 = R_t \mathbf{X}_0 + b_t \\ \lambda_1' \mathbf{x}_1 + \lambda_1 \mathbf{x}_1' = \widehat{w}_t R_t \mathbf{X}_0 + d_t \\ \lambda_2 \mathbf{x}_2 = S R_t \mathbf{X}_0 + S b_t + u \\ \lambda_2' \mathbf{x}_2 + \lambda_2 \mathbf{x}_2' = S \widehat{w}_t R_t \mathbf{X}_0 + S d_t + u \;\; . \end{cases} \tag{16}$$

These equations can be written as a single matrix equation:

$$\underbrace{\begin{bmatrix} R_t & \mathbf{x}_1 & 0 & 0 & 0 & b_t \\ \widehat{w}_t R_t & \mathbf{x}_1' & \mathbf{x}_1 & 0 & 0 & d_t \\ S R_t & 0 & 0 & \mathbf{x}_1 & 0 & S b_t + u \\ S \widehat{w}_t R_t & 0 & 0 & \mathbf{x}_2' & \mathbf{x}_2 & S d_t \end{bmatrix}}_{S_c} \begin{bmatrix} \mathbf{X}_0 \\ -\lambda_1 \\ -\lambda_1' \\ -\lambda_2 \\ -\lambda_2' \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \;\; , \tag{17}$$

implying $\operatorname{rank} S_c < 6$. By a suitable choice of object coordinates we may assume that $R_t = I$ and $b_t = 0$ – we may even assume that $S = I$ by a suitable change of coordinates in the second image and renaming u – and thus obtain

$$S_c = \begin{bmatrix} I & \mathbf{x}_1 & 0 & 0 & 0 & 0 \\ \widehat{w}_t & \mathbf{x}_1' & \mathbf{x}_1 & 0 & 0 & d_t \\ S & 0 & 0 & \mathbf{x}_1 & 0 & u \\ S \widehat{w}_t & 0 & 0 & \mathbf{x}_2' & \mathbf{x}_2 & S d_t \end{bmatrix} \;\; . \tag{18}$$

Again, the same constraint can be derived from a discrete differential starting point. The constraints within $\operatorname{rank} S_c < 6$ will be called *the hybrid stereo constraints*.

Expanding the minors of (18) gives the following different constraints in the motion parameters w_t and b_t:

1. Minors containing only one row out of rows 4 to 6 give the previously derived hybrid epipolar constraint between views 1 and 2.
2. Minors containing only one row out of the last three rows give the previously derived hybrid epipolar constraint between views 2 and 1.
3. Minors containing at least two rows out of row 4 to 6 and at least two rows out of the last three rows, give either trivial constraints or non-linear constraints in the parameters $\widehat{w}_t$ and d_t needed for the motion update.

The first and second type of hybrid stereo constraints above can be used to update a current structure and motion estimate linearly. I turns out that these constraints give two linearly independent constraints on the motion parameters, which again can be shown in the same manner as in Appendix A. This implies that the motion parameters of a rigidly moving stereo-rig can be updated from at least three corresponding points.

3.4 Case IV: 2D-cameras

In the case of 2D-cameras, the camera matrix equations (1), (2) and (3) looks exactly the same, but the camera matrix P is a 2×3-matrix, the orthogonal matrix R is a 2×2-matrix and the translational vector b is a 2-vector. Writing down the camera matrix equations for time 0, time t and time $t + \Delta t$, results in a matrix equation that looks exactly as (10), with the main difference that the size of M_d is now 4×4 and $\widehat{w}_t$ now denotes a 2×2 skew-symmetric matrix, i.e. a matrix defined by a single parameter w_t. The condition

$$\det \begin{bmatrix} R_t \mathbf{x}_0 & \mathbf{x}_t & 0 & b_t \\ \widehat{w}_t R_t \mathbf{x}_0 & \mathbf{u}_t & \mathbf{x}_t & d_t \end{bmatrix} = 0 \tag{19}$$

is called the *2D hybrid epipolar constraint*. This constraint consists of one single linear constraint on the three motion parameters w_t and d_t. Thus it is possible to update a current motion estimate from only three corresponding points in two images. This is a major improvement to standard methods in several ways. Firstly, the update equations are linear in the motion parameters. Secondly, only two images (along with the motion of the feature points in one of the images) are needed, compared to the traditional methods where at least three images are needed in order to obtain any constraints on the motion parameters. Finally, a unique update is obtained, compared to the discrete case where three images always gives an ambiguous solution, cf. [17].

Writing down the camera matrix equations for time 0, s, t and $t + \Delta t$, results in a matrix equation that looks exactly as (13), with the main difference that the size of M_d is now 6×5 and again $\widehat{w}_t$ now denotes a 2×2 skew-symmetric matrix, defined by a single parameter w_t. The condition

$$\operatorname{rank} \begin{bmatrix} R_s \mathbf{x}_0 & \mathbf{x}_s & 0 & 0 & b_s \\ R_t \mathbf{x}_0 & 0 & \mathbf{x}_t & 0 & b_t \\ \widehat{w}_t R_t \mathbf{x}_0 & 0 & \mathbf{u}_t & \mathbf{x}_t & d_t \end{bmatrix} \leq 5 \tag{20}$$

is called the *2D hybrid trifocal constraint*. Expanding the minors of (20) gives the following different constraints in the motion parameters w_t and b_t:

1. Minors containing only one row out of rows 1 and 2 give the previously derived 2D hybrid epipolar constraint between views 2 and 3.
2. Minors containing only one row out of rows 3 and 4 give linear constraints in the motion parameters, in total 2 such linear constraints.
3. Minors containing only one row out of rows 5 and 6 give the standard discrete trifocal constraint between views 1 and 2 and 3.

The second type of constraints are interesting for us. However, the two linear constraints obtained in this way are linearly dependent, implying that at least three images are needed in order to linearly update the current motion estimate. For a detailed proof, see Appendix B.

3.5 Motion estimation using HMC

Given a current estimate of the motion parameters and a new image with at least three corresponding points, the motion parameters can be updated using a linear system of equations of the type

$$M \begin{bmatrix} w_t \\ d_t \end{bmatrix} = m \ , \tag{21}$$

where $M = M(\mathbf{x}_0^k, \mathbf{x}_t^k, \mathbf{u}_t^k, R_t, b_t)$ for $k = 1, \dots n, n \geq 3$ and $m = m(\mathbf{x}_0^k, \mathbf{x}_t^k, \mathbf{u}_t^k, R_t, b_t)$ are the functions that computes the coefficients of the hybrid matching constraint equations in Sec. 3.1 in the epipolar case and similarly for the other cases.

3.6 State estimation using the continuous-discrete EKF

Given the motion parameters it is possible to employ a number of algorithms for recursive structure recovery, e.g. a continuous-discrete extended Kalman (EKF) filter for the state estimation process [18], [19].

3.7 Motion estimation refinement by reprojection constraints

Given motion estimates R_t and b_t obtained using the HMC through (21), the measurement $\mathbf{x}_t$, and the 3-D estimate $\mathbf{X}$ from the EKF, 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 the reprojection constraint

$$\lambda_t^+ \mathbf{x}_t = [\ R_t^+ \mid -b_t^+\ ]\mathbf{X} \ , \quad R_t^+ = e^{\widehat{\alpha}} R_t \ , \quad b_t^+ = b_t + \beta \ . \tag{22}$$

Expanding the first equation in (22) to a first order approximation gives

$$\begin{gathered} \lambda_t \mathbf{x}_t \approx R_t \tilde{\mathbf{X}} + \widehat{\alpha} R_t \tilde{\mathbf{X}} + b_t + \beta \ \Rightarrow \\ \widehat{R_t \tilde{\mathbf{X}}} \alpha + \beta = \epsilon := R_t \tilde{\mathbf{X}} + b_t - \lambda_t \mathbf{x}_t \ , \end{gathered} \tag{23}$$

where ϵ can be interpreted as the reprojection error. Observe that (23) is a linear constraint in the correction vectors α and β. Since (23) contains two linear constraints on these 6 parameters (λ_t is also a free parameter) in the correction vectors, a linear update on the motion parameters can be made from at least three corresponding points. Similar correction formulas can be derived for the 2D-case. Observe that this motion estimation refinement is in fact a pose update, similar to one iteration in a non-linear least squares estimation problem.

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.

3.8 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. Initialization

- Assume that images are obtained sequentially at time instants t_i, $i = 0, 1, 2 \ldots$, equally spaced by Δt. Also assume some initial values for the state vector and the error covariance matrix in the EKF.
- 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 ν_0 using e.g. the continuous eight-point algorithm in the 3D single camera case.
- Compute $R_{t_1} = e^{\widehat{w}_0 \Delta t}$ and $b_{t_1} = d_0 \Delta t$.

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

- Using at least three point correspondences, set up the hybrid matching constraints in (21).
- Solve the linear system (21) for the new parameter estimates w_{t_i} and d_{t_i}.
- Update the rotation matrix and the translation vector according to (11).
- Use w_{t_i} and ν_{t_i} in the EKF to get structure estimates over the time interval $[t_i, t_i + \Delta t]$.
- Refine the motion estimate according to (23).

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 ν 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 (21).

4 Experiments

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 ν are time varying. The hybrid-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 images of an object consisting of eight points in a general configuration on a grid of stepsize 10^{-4}, and with the parameter vectors

$$
\begin{aligned}
w(t) &= \frac{2}{3}\begin{pmatrix}1, & -1, & 1\end{pmatrix} , \\
\nu(t) &= \frac{1}{\sqrt{1.32}}\begin{pmatrix}-1, & -0.4, & 0.4\end{pmatrix}^T + \frac{1}{2}\begin{pmatrix}t, & t, & -0.5t\end{pmatrix}^T .
\end{aligned}
$$

Perspective measurements were computed at time instants separated by $\Delta t = 0.01$. The estimates of the components of the rotational velocity w and the translational velocity ν together with the true values based on the hybrid epipolar constraints, are shown in

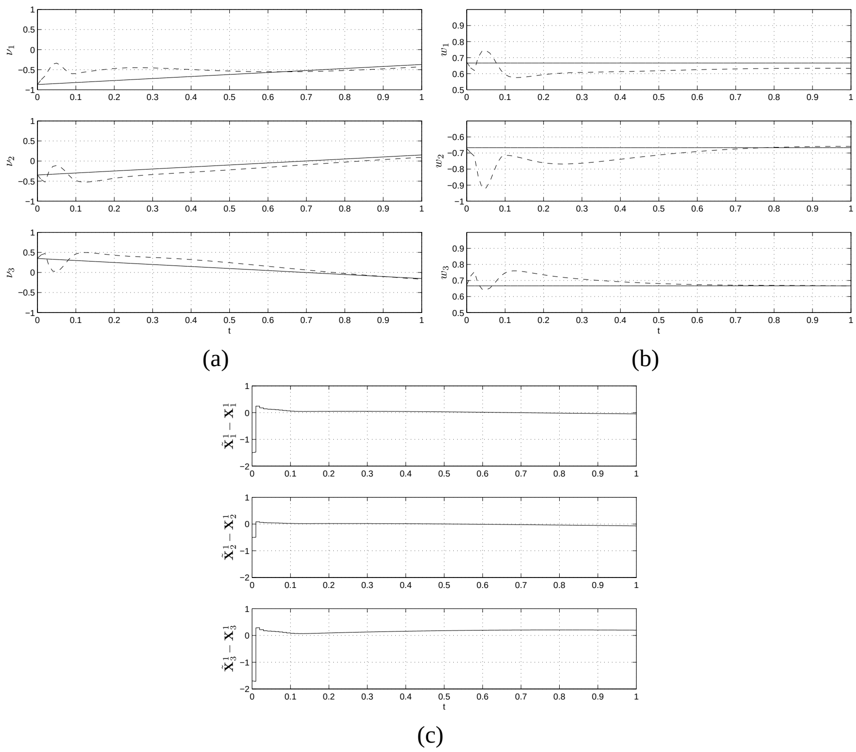


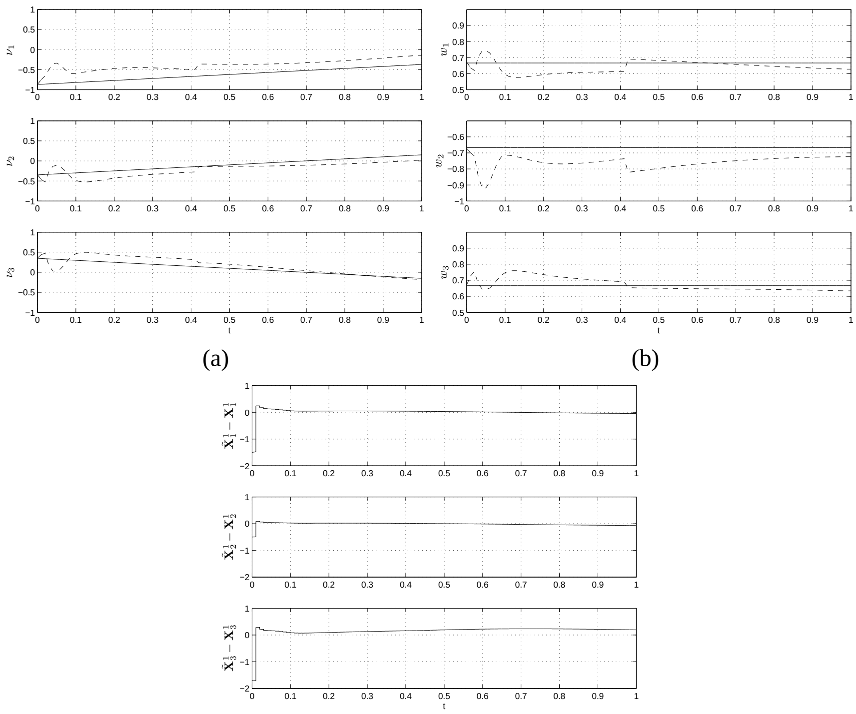
Fig. 1. Estimation results obtained using the hybrid epipolar constraints: (a) True (solid) and estimated (dashed) translational velocity ν, (b) True (solid) and estimated (dashed) rotational velocity w, (c) 3-D estimation errors for one of the observed object points.

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. $1(c)$.

We also conducted a similar experiment on the same data based on the hybrid trifocal constraints. The same procedure as before, based on the initialization using the continuous epipolar constraint, and the recursive estimation using the epipolar hybrid constraints, was used until time $t = 0.4s$. After that, the trifocal hybrid constraints was used, with $s = t/2$, see Fig 2.

5 Conclusion

We have derived several hybrid matching constraints; epipolar-, trifocal-, stereo- and 2D-hybrid motion constraints. We have proposed an algorithm for on-line recursive estimation of structure and motion from perspective measurements in a continuous-discrete setting, utilizing these hybrid matching constraints 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 processes are connected



(c)

Fig. 2. Estimation results obtained using the hybrid trifocal constraints: (a) True (solid) and estimated (dashed) translational velocity ν, (b) True (solid) and estimated (dashed) rotational velocity w, (c) 3-D estimation errors for one of the observed object points.

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. The main advantages of the presented method is that only three corresponding points are needed for the sequential update and correction of the velocity parameter estimates. Further, both these update schemes are linear.

Note that it is *not* necessary that the same three points are tracked throughout the whole image sequence. It is easy to change to any other triplet of point correspondences when needed.

References

1. Hartley, R., Zisserman, A.: Multiple View Geometry. Cambridge (2003)
2. Longuet-Higgins, H.C.: A computer algorithm for reconstructing a scene from two projections. Nature **293** (1981) 133–135
3. Torr, P., Fitzgibbon, A.W., Zisserman, A.: Maintaining multiple motion model hypotheses over many views to recover matching and structure. In: Proc. International Conference on Computer Vision. (1998) 485–491

4. Åström, K., Heyden, A.: Continuous time matching constraints for image streams. International Journal of Computer Vision **28**(1) (1998) 85–96
5. Viéville, T., Faugeras, O.: The first order expansion of motion equations in the uncalibrated case. Computer Vision and Image Understanding **64**(1) (1996) 128–146
6. Ma, Y., Košecká, J., Sastry, S.S.: Linear differential algorithm for motion recovery: A geometric approach. International Journal of Computer Vision **36**(1) (2000) 71–89
7. Triggs, B.: Differential matching constraints. In: Proc. of the International Conference on Computer Vision and Pattern Recognition. Volume 1. (1999) 370–376
8. Heyden, A.: Differential-algebraic multiview constraints. In: Proc. International Conference on Pattern Recognition. Volume 1., IEEE Computer Society Press (2006) 159–162
9. Nyberg, F., Heyden, A.: Recursive structure from motion using hybrid matching constraints with error feedback. In: ICCV 2005 and ECCV 2006 Workshops, WDV 2005 and WDV 2006. Volume 4358. (2007) Springer Lecture Notes in Computer Science.
10. Soatto, S.: 3-D structure from visual motion: Modeling, representation and observability. Automatica **33**(7) (1997) 1287–1312
11. Azarbayejani, A., Pentland, A.P.: Recursive estimation of motion, structure, and focal length. IEEE Transactions on Pattern Analysis and Machine Intelligence **17**(6) (195) 562–575
12. Hyyppä, K.: Optical navigation system using passive identical beacons. In: Proc. Intelligent Autonomous Systems. (1997)
13. Quan, L., Kanade, T.: Affine structure from line correspondences with uncalibrated affine cameras. IEEE Trans. Pattern Analysis and Machine Intelligence **19**(8) (1997) 834–845
14. Triggs, B.: Matching constraints and the joint image. In: Proc. International Conference on Computer Vision. (1995) 338–343
15. Heyden, A.: Tensorial properties of multilinear constraints. Mathematical Methods in the Applied Sciences **23** (2000) 169–202
16. Faugeras, O., Luong, Q.T.: Geometry of Multiple Images. The MIT Press (2001)
17. Åström, K., Kahl, F.: Ambiguous configurations for the 1d structure and motion problem. Journal of Mathematical Imaging and Vision **18**(2) (2003) 191–203
18. Maybeck, P.S.: Stochastic Models, Estimation, and Control Volume 2. Academic Press (1982)
19. Huster, A.: Relative Position Sensing by Fusing Monocular Vision and Inertial Rate Sensors. PhD dissertation, Stanford University, Department of Electrical Engineering (2003)

Appendix A

In order to prove that there exist exactly two linearly independent constraints among the nine epipolar hybrid constraints, we have to prove that there are exactly seven independent linear dependencies between these constraints. Start by extending M_d in (10) with the second column, giving the following 6×5-matrix:

$$\begin{bmatrix} R_t\mathbf{x}_0 & \mathbf{x}_t & 0 & b_t & \mathbf{x}_t \\ \widehat{w}_t R_t \mathbf{x}_0 & \mathbf{u}_t & \mathbf{x}_t & d_t & \mathbf{u}_t \end{bmatrix} \ .$$

The 5×5 minors of this matrix are obviously identically zero. Expanding the three minors obtained by removing each one of the last three columns give

$$\begin{aligned} &x_t L_{11} - y_t L_{21} + L_{31} - u_x y_t E + u_y x_t E = 0 \\ &x_t L_{12} - y_t L_{22} + L_{32} - u_x E = 0 \\ &x_t L_{13} - y_t L_{23} + L_{33} + u_y E = 0 \ , \end{aligned} \tag{24}$$

where L_{ij} denotes the hybrid constraint obtained by removing row number i from the first three rows of M_d and row number j from the last three rows of M_d and E denotes the epipolar constraint between time 0 and t. Thus (24) contains exactly three linear dependencies among the epipolar hybrid constraints, assuming that the epipolar constraint between views 0 and t is fulfilled. Repeating the same procedure by instead adding the third column and expanding the minors obtained by removing each one of the first three columns give three further linear dependencies. Finally, adding both the second and the third column and expanding the determinant of the resulting 6×6 matrix, give the last linear dependency. It is furthermore evident from the structure of these seven dependencies that they are linearly independent. Thus only two linearly independent hybrid constraints remain (assuming the epipolar constraint is fulfilled). For practical purposes, there might be more than two linearly independent constraints when the epipolar constraint is not exactly fulfilled. However, these are numerically ill-conditioned in the sense that they are close to spanning a two-dimensional space.

Appendix B

In order to prove that the two 2D trifocal hybrid constraints are linearly dependent we can proceed similarly as in Appendix A. Start by extending the matrix in (20) with the third column, giving the following 6×6-matrix:

$$\begin{bmatrix} R_s\mathbf{x}_0 & \mathbf{x}_s & 0 & 0 & b_s & 0 \\ R_t\mathbf{x}_0 & 0 & \mathbf{x}_t & 0 & b_t & \mathbf{x}_t \\ \widehat{w}_t R_t\mathbf{x}_0 & 0 & \mathbf{u}_t & \mathbf{x}_t & d_t & \mathbf{u}_t \end{bmatrix}$$

. This 6×6 matrix is obviously singular, with determinant identically zero. This determinant can be written as

$$x_t L_1 - y_t L_2 + u_x u_y T - u_y u_x T = x_t L_1 - y_t L_2 = 0 \ , \tag{25}$$

where L_i denotes the 2D hybrid epipolar constraints and T denotes the discrete trifocal constraint, giving the desired linear dependency between the two hybrid epipolar constraints.