

# Using a Connected Filter for Structure Estimation in Perspective Systems<sup>†</sup>

Fredrik Nyberg, Ola Dahl, Jan Holst and Anders Heyden

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

## Abstract

*Three-dimensional structure information can be estimated from two-dimensional images using recursive estimation methods. This paper investigates possibilities to improve structure filter performance for a certain class of stochastic perspective systems by utilizing mutual information, in particular when each observed point on a rigid object is affected by the same process noise. After presenting the dynamic system of interest, the method is applied, using an extended Kalman filter for the estimation, to a simulated time-varying multiple point vision system. The performance of a connected filter is compared, using Monte Carlo methods, to that of a set of independent filters. The idea is then further illustrated and analyzed by means of a simple linear system. Finally more formal stochastic differential equation aspects, especially the impact of transformations in the Itô sense, are discussed and related to physically realistic noise models in vision systems.*

## 1. Introduction

In many computer vision applications it is required that three-dimensional information can be estimated from two-dimensional images. It is possible to estimate three-dimensional parameters related to structure and/or motion, given a sequence of images, by employing an algorithm that utilizes information from all images in the sequence as input data. An overview of this type of algorithms can be found in e.g. [7], [11]. Another class of algorithms uses 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 algorithm typically performs *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 [19]. Estimation of *structure* using dynamic systems is described e.g. in [18], which contains results regarding observability, and also presents algorithms and experimental results.

The algorithms presented in [18] and [19] describe structure and motion estimation using a camera as the only measurement device. Algorithms can also be developed for situations where additional measurement devices are used, or where additional knowledge of specific quantities are available, e.g. situations where the linear and/or angular velocity of an object is measured or estimated. This type of algorithms includes e.g. estimation of orientation, using a camera and an inertial measurement unit, as shown in [17], or estimation of orientation and position, using a camera moving with known angular and linear velocities, as presented in [1].

Recursive estimation of the three-dimensional position of one or more feature points on a rigid object, assuming measurements or estimated values of the angular and the linear velocities, can be approached using an extended Kalman filter [2], [12], or other types of more specialized nonlinear observers [3]–[5], [10], [13].

The observers in [3]–[5], [10], [13] estimate the position of a single feature point. This paper investigates performance issues in the simultaneous position estimation of multiple points. An extended Kalman filter is used for the estimation. It is exemplified how the estimation can be improved using a connected filter, under the assumption of a special process noise structure.

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

## 2. Perspective system

The object motion is described using a dynamic perspective system. The perspective system combines a set of stochastic differential equations for the position of a number of feature points on an observed object, with noise corrupted measurement equations in the form of perspective projections defined using an appropriate camera model. In this section the corresponding deterministic dynamic system is presented under the assumptions of rigid body motion and a simple frontal pinhole imaging model. A coordinate transformation is employed to obtain a useful alternative system formulation. The mathematical model is then extended to include process and measurement noise, and consequently re-formulated in stochastic differential form.

### 2.1. Dynamic system

The three-dimensional position of an arbitrarily selected feature point on a given object is described using coordinates  $x_i = (x_{i,1} \ x_{i,2} \ x_{i,3})^T$ ,  $i = 1 \dots n$ , where  $n$  is the number of points. Under the assumption of rigid body motion,  $x_i$  satisfies the differential equation

$$\dot{x}_i = Ax_i + b, \quad x_i(0) = x_{0i}, \quad (1)$$

where  $A$  is the skew-symmetric matrix

$$A = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix},$$

parameterized by the angular velocity vector  $\omega = (\omega_1 \ \omega_2 \ \omega_3)^T$ , and  $b$  is the vector

$$b = (b_1 \ b_2 \ b_3)^T = \dot{d} - Ad,$$

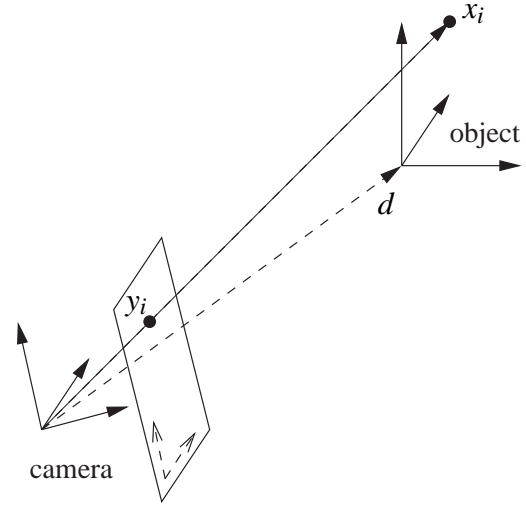
where  $d$  denotes the vector from the origin of an inertial coordinate system to the origin of a local coordinate system, attached to the body, as illustrated in Fig. 1. For a detailed description of the characteristics of rigid body motion, see e.g. [15].

A camera model is defined as a transformation of the three-dimensional coordinates of a feature point  $x_i$ , resulting in projected image coordinates  $y_i$  for the point. Considering an object point with three-dimensional coordinates expressed by the vector  $x_i$ , and using a frontal pinhole imaging model [11], where the optical axis is chosen to coincide with the direction of  $x_{i,3}$ , the following transformation rule can be derived,

$$y_i = \frac{1}{x_{i,3}} \begin{pmatrix} m_1 f & s & 0 \\ 0 & m_2 f & 0 \end{pmatrix} x_i = \frac{1}{x_{i,3}} Cx_i, \quad (2)$$

where the parameter  $f$  is the focal length,  $s \geq 0$  is the skew factor, describing the situation where the image coordinate

axes are not perpendicular, and  $m_1$  and  $m_2$  are strictly positive scaling factors, acting in the horizontal and vertical direction of the image respectively.



**Fig. 1. Coordinate systems for the position of a three-dimensional point  $x_i$ , belonging to an observed object, and its projection  $y_i$  onto an image plane.**

Combining (1) and (2), a dynamic system with state vectors  $x_i$ ,  $i = 1 \dots n$  and corresponding output vectors  $y_i = (y_{i,1} \ y_{i,2})^T$  can be constructed as

$$\begin{aligned} \dot{x}_i &= Ax_i + b, \quad x_i(0) = x_{0i} \\ y_i &= \frac{1}{x_{i,3}} Cx_i. \end{aligned} \quad (3)$$

Since the system (3) describes the motion of three-dimensional feature points observed by images obtained under perspective projection, it is usually denoted a *perspective system* [6], or sometimes more specifically, since the state equations are linear in the states, a *linear perspective system* [9].

In the case of rigid body motion, the parameter matrices  $A$  and  $b$  in (3) are usually time-varying. Also the internal camera parameters in  $C$  could change with time, e.g. as a consequence of zooming. Equation (3) then constitutes an affine, time-varying, dynamic system for the three-dimensional state  $x_i$ , combined with a time-varying nonlinear output equation. For the sake of clarity of presentation, the time-dependency of  $A$ ,  $b$  and  $C$  will in the following in general not be made explicit.

For the purpose of this paper the camera is assumed to be calibrated and the motion parameters given, either by construction or by measurements, i.e. the matrices  $A$ ,  $b$  and  $C$  are assumed known.

## 2.2. Coordinate transformation and stochastic generalizations

Introduce the coordinate transformation

$$z_{i,1} = \frac{x_{i,1}}{x_{i,3}}, \quad z_{i,2} = \frac{x_{i,2}}{x_{i,3}}, \quad z_{i,3} = \frac{1}{x_{i,3}}, \quad (4)$$

which is used also in e.g. [3]–[5], [10]. Also define the new parameter matrices

$$\begin{aligned} \bar{A} &= \begin{pmatrix} 0 & -\omega_3 & b_1 \\ \omega_3 & 0 & b_2 \\ 0 & 0 & 0 \end{pmatrix} \\ \bar{b} &= (\omega_2 \quad -\omega_1 \quad 0)^T \\ \bar{c} &= (\omega_2 \quad -\omega_1 \quad -b_3)^T. \end{aligned} \quad (5)$$

The system (3) can then be transformed to a new system in the  $z$ -coordinates as

$$\begin{aligned} \dot{z}_i &= \bar{A}z_i + \bar{b} + (\bar{c}^T z_i)z_i, \quad z_i(0) = z_{0i} \\ y_i &= Cz_i. \end{aligned} \quad (6)$$

Note that due to the transformation, the output equation is now linear in the states, and that the nonlinearities have been shifted to the state equations.

An important issue in many computer vision-related estimation applications is the presence of noise inputs, due to e.g. motion inaccuracies and measurement disturbances, as well as uncertainties in the mathematical model itself. To obtain more general models, it is therefore desirable to include stochastic elements in the dynamic system equations. This can be done in various ways, and here we consider the following generalizations of (3) and (6), represented by stochastic integral equations written in differential form as

$$\begin{aligned} dx_i &= (Ax_i + b)dt + G_x(x_i, t)d\beta_{wi}, \quad x_i(0) = x_{0i} \\ d\bar{y}_i &= \frac{1}{x_{i,3}}Cx_i dt + Dd\beta_{vi}, \end{aligned} \quad (7)$$

and

$$\begin{aligned} dz_i &= [\bar{A}z_i + \bar{b} + (\bar{c}^T z_i)z_i]dt + G_z(z_i, t)d\beta_{wi}, \quad z_i(0) = z_{0i} \\ d\bar{y}_i &= Cz_i dt + Dd\beta_{vi}, \end{aligned} \quad (8)$$

for  $i = 1 \dots n$ , where the integrated measurement representations

$$\bar{y}_i(t) = \int_0^t y_i(\tau) d\tau$$

have been introduced. The noise influence is described using mutually independent vector-valued standard Wiener processes  $\beta_{wi}$  and  $\beta_{vi}$ , two functions  $G_x$  and  $G_z$ , and a matrix  $D$ .  $\beta_{wi}$  and  $\beta_{vi}$  are referred to as *process noise* and *measurement noise* respectively. A detailed treatment of the formalism and interpretation of stochastic differential equations on the form (6) can be found in [16].

## 3. Filter algorithm

A structure from motion filter design problem can now be formulated as the task of recursively estimating  $n$  unknown states  $x_i$ , each governed by the system model (7), at the time  $t$ , given the corresponding perspective measurements  $y_i$  up to that instant, in a way that is in some sense optimal. Another possibility is to use the transformed formulation (8) for the estimation, and then recover the three-dimensional position by an inverse transformation.

In designing a filter for either of the systems (7) and (8) it is possible to consider the motion of each point as a separate 3-state system, and apply to each an individual filter. Thus the state estimation is performed using *independent filters* for the  $n$  points. In this approach obviously no improvement in the estimate for any individual point is made compared to the single point case. Another approach is to assemble the  $n$  feature point equations into a single  $3n$ -dimensional system, and construct a *connected filter* which can utilize any common information present. For example, it is a reasonable conjecture that if there exists some kind of correlation or other dependency between the noise input vectors, the connected filter approach rightly handled will lead to a better estimation.

### 3.1. The extended Kalman filter

State estimation in a system of equations can be performed in various ways. A widely used technique for nonlinear filtering is the extended Kalman filter (EKF) [14].

Consider a nonlinear dynamic system in the rather general form

$$\begin{aligned} dz &= f(z, t)dt + Bd\beta_w \\ d\bar{y} &= h(z, t)dt + Dd\beta_v. \end{aligned} \quad (9)$$

For simplicity of notation, introduce the diffusion strength matrices  $Q$  and  $R$ , such that  $Q \doteq BB^T$  and  $R \doteq DD^T$ . It is assumed in the following that  $B$  and  $D$  are known, and that  $R$  is positive definite. Also let  $F(\hat{z}, t)$  and  $H(\hat{z}, t)$  denote the partial derivative matrices of  $f$  and  $h$ , respectively, with respect to  $z$ , evaluated at  $z = \hat{z}$ . In the EKF, the estimate update equation is then given by

$$\dot{\hat{z}} = f(\hat{z}, t) + K(\bar{y} - h(\hat{z}, t)), \quad \hat{z}(0) = \hat{z}_0, \quad (10)$$

with the gain  $K$  computed as

$$K = PH^T(\hat{z}, t)R^{-1},$$

and  $P$  determined by the continuous time matrix Riccati differential equation

$$\begin{aligned} \dot{P} &= F(\hat{z}, t)P + PF^T(\hat{z}, t) + Q - PH^T(\hat{z}, t)R^{-1}H(\hat{z}, t)P \\ P(0) &= P_0. \end{aligned} \quad (11)$$

Note that this formulation of the EKF assumes that measurements are continuously available, while in many applications, especially in computer vision, it is more natural to consider the outputs as being given at discrete time points. Details on the EKF algorithm for the case of discrete time measurements can be found e.g. in [14].

It has been observed in the literature, see e.g. [8], that the standard filters, such as the EKF, in general performs poorly when applied directly to perspective-type systems. It has also been demonstrated by simulations and experiments that by applying a coordinate transformation of the type (4), that linearizes the measurement equations and shifts the nonlinearities to the state equations, these adverse effects can be significantly reduced [8]. It is therefore motivated to perform the structure from motion filtering in the system (8).

Now from (8) we obtain the following functions corresponding to those present in (10) and (11)

$$\begin{aligned} f(\hat{z}_i, t) &= \bar{A} + \bar{b} + (\bar{c}^T \hat{z}_i) \hat{z}_i, & F(\hat{z}, t) &= \bar{A} + (\bar{c}^T \hat{z}) I + \hat{z} \bar{c}^T, \\ h(\hat{z}_i, t) &= C \hat{z}_i, & H(\hat{z}, t) &= C. \end{aligned}$$

### 3.2. Observing multiple points with identical process noise

Assume that  $n$  points are being observed under perspective projection. Further make the assumption that the motion of each individual point  $x_i$  is such that it can be described in the transformed coordinates  $z_i$  by a dynamic system of the type (8). As mentioned above, it is expected that the estimation performance improvement when using a connected filter will be more significant when the process noise is correlated. As an extreme case, assume that the process noise vectors influencing each coordinate  $z_i$  are *identical*, i.e.  $\beta_{wi} = \beta_w$  for  $i = 1 \dots n$ . The measurement noise vectors  $\beta_{vi}$  on the other hand, are assumed to be different for different points. The systems can then be described in the transformed coordinates as

$$\begin{aligned} dz_i &= [\bar{A}z_i + \bar{b} + (\bar{c}^T z_i) z_i] dt + B d\beta_w, & z_i(0) &= z_{0i} \\ d\bar{y}_i &= C z_i dt + D_i d\beta_{vi}, \end{aligned} \quad (12)$$

for  $i = 1 \dots n$ , where  $B$  is assumed to be a constant matrix.

Define  $\mathbf{z}$ ,  $\bar{\mathbf{y}}$  and  $\beta_{\mathbf{v}}$  to be the vectors obtained by stacking the  $n$  vectors  $z_i$ ,  $\bar{y}_i$  and  $\beta_{vi}$  respectively, and  $\bar{\mathbf{b}}$  the vector obtained by stacking  $n$  copies of  $\bar{b}$ . Further let  $\bar{\mathbf{A}}$  and  $\mathbf{C}$  be the matrices obtained by tiling  $n$  copies of  $\bar{A}$  and  $C$  into diagonal matrices. Similarly  $\bar{\mathbf{c}}$  is defined to be the  $3n \times 3n$  block diagonal matrix with diagonal elements  $\bar{c}^T z_i I_{3 \times 3}$ , and  $\mathbf{D}$  the  $2n \times 2n$ -dimensional block diagonal matrix with diagonal elements  $D_i$ . The system (12) can then be written as

$$\begin{aligned} d\mathbf{z} &= [\bar{\mathbf{A}}\mathbf{z} + \bar{\mathbf{b}} + \bar{\mathbf{c}}\mathbf{z}] dt + \mathbf{B}d\beta_w, & \mathbf{z}(0) &= \mathbf{z}_0 \\ d\bar{\mathbf{y}} &= \mathbf{C}z dt + \mathbf{D}d\beta_{\mathbf{v}}, \end{aligned} \quad (13)$$

where  $\mathbf{B}$  and  $\beta_w$  are chosen as follows. Applying  $n$  independent filters to the system (13) is equivalent to choosing  $\mathbf{B}$  to be a block diagonal matrix with  $n$  copies of  $B$  as its diagonal elements and zeros elsewhere, i.e.

$$\mathbf{B} = \begin{pmatrix} B & & & \\ & B & & \\ & & \ddots & \\ & & & B \end{pmatrix}. \quad (14)$$

To keep dimensions consistent,  $\beta_w$  is then defined to be the vector obtained by stacking  $n$  copies of  $\beta_w$ . On the other hand, a connected filter with  $n$  points is obtained by selecting  $\mathbf{B}$  as the matrix

$$\mathbf{B} = (B^T \quad B^T \quad \dots \quad B^T)^T, \quad (15)$$

where  $n$  copies of  $B$  have been used, and defining  $\beta_w$  to be equal to  $\beta_w$ .

The choice of structure for  $\mathbf{B}$  determines the structure of  $Q$  in (11), and it can be seen by inspection that, given a diagonal initial matrix  $P_0$ , a block diagonal  $\mathbf{B}$  as in (14), will yield a block diagonal solution  $P(t)$ , whereas a choice according to (15) will lead to a full solution matrix  $P(t)$ .

### 3.3. A perspective system example

This section illustrates the improvement gained by utilizing the proposed connected filter design on a stochastic perspective system with time-varying parameters. All simulations were done using Matlab.

Consider the transformed stochastic perspective system (12) governed by the parameter vectors and matrices

$$\begin{aligned} \omega &= \pi (0.4 \quad 0.5 \quad 0.5) \\ b &= (1 \quad -0.5t + 0.4 \quad 0.5t - 0.4)^T \\ B &= I_{3 \times 3}, \quad D = 0.01 \cdot I_{2 \times 2}, \end{aligned}$$

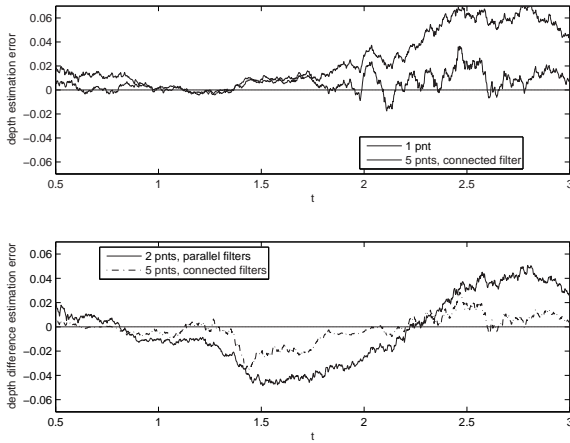
and with the perspective projection parameters in (2) given by  $f = m_1 = m_2 = 1$  and  $s = 0$ . The system (12) was simulated by employing an explicit fixed-step second order differential equation solver on a grid of step size 0.001. The effect of the noise terms was approximated by addition of pseudo-random normal distributed numbers of zero mean and the proper variance, in each step of the integration. To initiate the Riccati differential equation (11) employed in the EKF filter, the initial covariance matrix was set to  $P_0 = 300 \cdot I_{3 \times 3}$ . The filter was implemented using the Dormand-Prince algorithm.

In order to investigate the influence of a connected filter on this system a hundred-run Monte Carlo simulation was performed. It included estimation using both independent filters and a connected filter for up to five three-dimensional points for each noise realization. For each realization identical process noise was injected, as in (12),

to each three-dimensional transformed state vector  $z_i$ . The results for one of these points are shown for comparison in Fig. 2 and Fig. 3. Let a normalized variance  $V_0$  be defined as

$$V_0 = \frac{\text{var}(x_{1,3} - \hat{x}_{1,3}^c)}{\text{var}(x_{1,3} - \hat{x}_{1,3}^p)},$$

where superscripts  $c$  and  $p$  denotes the estimates obtained using the connected filter and the independent filters respectively. For the difference estimation error plot, shown in the lower plot in Fig. 3,  $V_0$  was similarly constructed. In both cases,  $V_0$  was computed based on the data from the same time interval as that shown in Fig. 2. It can

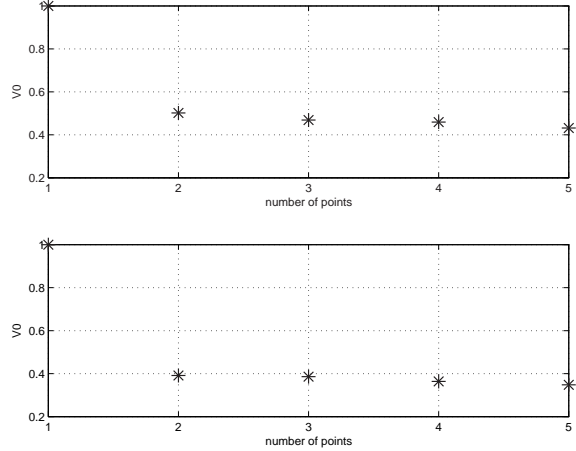


**Fig. 2. Depth estimation error  $x_{1,3} - \hat{x}_{1,3}$  (top), and depth difference estimation error  $(x_{1,3} - x_{2,3}) - (\hat{x}_{1,3} - \hat{x}_{2,3})$ , averaged over 100 different noise realizations.**

be seen in the upper plot in Fig. 2 that the connected filter tends to center the estimation error around zero, thus decreasing the drift seen in the independent filter estimation. In Fig. 3 it is clearly seen that utilizing several points and a connected filter in the estimation process leads to a significantly decreased variance for the estimation error, compared to using independent filters. It can also be seen that the decrease in the *difference* estimation error variance is even more distinct when connecting two or more filters. On the other hand the differences in error variances when comparing a connected filter with two or more points are more modest.

#### 4. Explicit analysis of a linear example

This section is intended to further study and exemplify the idea that if a connected filter is used, the state estimation results can be improved for a certain class of noise models.



**Fig. 3. Normalized variance for the depth estimation error  $x_{1,3} - \hat{x}_{1,3}$  (top), and for the depth difference estimation error  $(x_{1,3} - x_{2,3}) - (\hat{x}_{1,3} - \hat{x}_{2,3})$ .**

#### 4.1. Analytical variance comparison for single state systems

For the purpose of illustration, consider two one-dimensional linear systems, with corresponding linear output equations, affected by the same process noise, according to

$$\begin{aligned} dz_1 &= az_1 dt + \sqrt{\sigma} d\beta_w, & d\bar{y}_1 &= z_1 dt + \sqrt{\eta} d\beta_{v1} \\ dz_2 &= az_2 dt + \sqrt{\sigma} d\beta_w, & d\bar{y}_2 &= z_2 dt + \sqrt{\eta} d\beta_{v2} \end{aligned} \quad (16)$$

where as before  $\beta_w$  and  $\beta_{vi}$ ,  $i = 1, 2$ , are mutually independent vector-valued standard Wiener processes. The parameter  $a$  is assumed to be a real constant, and  $\eta$  and  $\sigma$  are assumed to be strictly positive. Introducing the system matrices

$$F = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and assuming stationarity, i.e.  $\dot{P} = 0$ , the Riccati differential equation becomes

$$P^2 - 2a\eta P - \eta\sigma BB^T = 0. \quad (17)$$

Here the structure of the matrix  $B$  determines the filter design philosophy. Equation (17) is a second order matrix polynomial equation, with solutions

$$P = a\eta I \pm \sqrt{a^2\eta^2 I + \eta\sigma BB^T}. \quad (18)$$

The solution to (18) now splits into two cases, depending on the method chosen to construct the Kalman filter.

**Independent filters:** The filters for the two states

are run in parallel. This is accomplished by letting  $B = I$ , and thus, by (18),

$$P_p = a\eta I \pm \sqrt{a^2\eta^2 + \eta\sigma} I.$$

Since  $P$  is a covariance matrix and  $\eta, \sigma > 0$ , the minus sign yields an infeasible solution. Hence one obtains an expression for the covariance matrix for the independent filters as

$$P_p = a\eta I + \sqrt{a^2\eta^2 + \eta\sigma} I. \quad (19)$$

**Connected filter:** Here a connected filter is constructed to utilize the common noise information. This implies  $B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}^T$ , and hence

$$BB^T = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Consequently, by (18),

$$P_c = a\eta I \pm \sqrt{\begin{pmatrix} (a^2\eta^2 + \eta\sigma) & \eta\sigma \\ \eta\sigma & (a^2\eta^2 + \eta\sigma) \end{pmatrix}} \doteq a\eta I \pm \sqrt{W(a, \eta, \sigma)}.$$

Using the decomposition  $W = V^{-1}DV$ , with  $D$  diagonal (which always exists since  $W$  is symmetric) gives

$$P_c = a\eta I \pm \sqrt{V^{-1}DV} = a\eta I \pm V^{-1}\sqrt{DV}. \quad (20)$$

For this simple system it is easy to find analytical expressions for the decomposition matrices  $D$  and  $V$  as

$$D = \begin{pmatrix} a^2\eta^2 & 0 \\ 0 & a^2\eta^2 + 2\eta\sigma \end{pmatrix}, \quad V = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The minus sign in (20) clearly gives an infeasible solution, and the expression for the covariance matrix for the connected filter becomes

$$P_c = \frac{1}{2} \begin{pmatrix} (2a+|a|)\eta + \sqrt{a^2\eta^2 + 2\eta\sigma} & -|a|\eta + \sqrt{a^2\eta^2 + 2\eta\sigma} \\ -|a|\eta + \sqrt{a^2\eta^2 + 2\eta\sigma} & (2a+|a|)\eta + \sqrt{a^2\eta^2 + 2\eta\sigma} \end{pmatrix}.$$

Thus for  $a \geq 0$  (unstable systems),

$$P_{cu} = \frac{1}{2}a\eta \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} + \frac{1}{2}\sqrt{a^2\eta^2 + 2\eta\sigma} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \quad (21)$$

Similarly for  $a < 0$  (stable systems),

$$P_{cs} = \left[ \frac{1}{2}a\eta + \frac{1}{2}\sqrt{a^2\eta^2 + 2\eta\sigma} \right] \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \quad (22)$$

Denote by  $V_p$ ,  $V_{cu}$  and  $V_{cs}$  the estimation error variance for one of the states, i.e. one of the diagonal elements in (19), (21) and (22) respectively. By straightforward analysis it can be shown that  $V_p \geq V_{cu}$  for  $a \geq 0$  and  $\eta, \sigma > 0$ , and that  $V_p \geq V_{cs}$  for  $a < 0$  and  $\eta, \sigma > 0$ .

For the case  $a > 0$ , introducing

$$\kappa = \frac{\sigma}{a^2\eta},$$

results in

$$\frac{V_p}{V_{cu}} = \frac{3 + \sqrt{1 + 2\kappa}}{2 + 2\sqrt{1 + \kappa}}.$$

The inequality  $V_p \geq V_{cu}$  is thus proven if it can be shown that

$$g(\kappa) \doteq 2\sqrt{1 + \kappa} - \sqrt{1 + 2\kappa} \geq 1,$$

for  $\forall \kappa > 0$ . But

$$g(0) = 1$$

and

$$g(\kappa) \approx (2 - \sqrt{2})\sqrt{\kappa} > 1, \quad \kappa \gg 0.$$

It thus suffices to show that  $g(\kappa)$  is monotone. But this follows directly from the inequality

$$\frac{dg}{d\kappa} = \frac{1}{\sqrt{1 + \kappa}} - \frac{1}{\sqrt{1 + 2\kappa}} > 0.$$

Similarly it can be shown that  $V_p \geq V_{cs}$  for  $a < 0$  and  $\eta, \sigma > 0$ . Consequently the connected filter strategy results in a decrease in the variance of the estimation error.

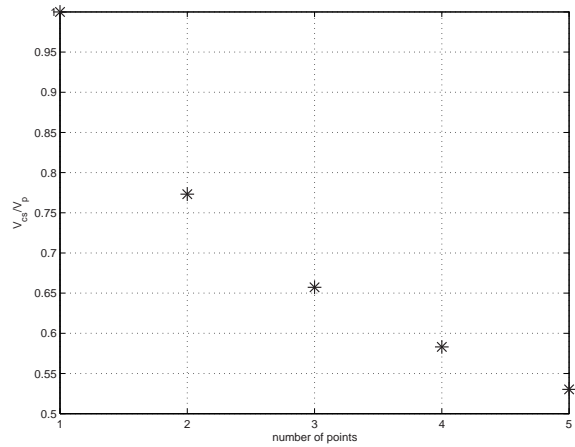
The above analysis can easily be extended to  $n$  one-dimensional systems, yielding for example for  $a < 0$

$$V_{cs} = \frac{1}{n}a\eta + \frac{1}{n}\sqrt{a^2\eta^2 + n\eta\sigma},$$

and

$$V_p = a\eta + \sqrt{a^2\eta^2 + \eta\sigma}.$$

In Fig. 4 the variance ratio  $V_{cs}/V_p$  is plotted against the number of points  $n$  for the parameter values  $a = -10$ ,  $\sigma = 10^{-2}$  and  $\eta = 10^{-4}$ . By comparison with the upper plot in Fig. 3 it can be seen that the behavior is similar.



**Fig. 4. Normalized variance for the linear system steady state estimation error  $z_1 - \hat{z}_1$  vs. the number of points utilized in the connected filter.**

By comparing the lower and upper plots in Fig. 3 it can be seen that the improvement in the difference

estimation error variance is slightly better than that for the depth estimation error variance. This type of improvement can be seen also in the linear example. For example, in the stable system case all elements of the estimation error covariance matrix are equal, as can be seen in (22). Therefore the stationary value of the variance of the difference in estimation errors is identically zero in this case.

## 5. More accurate noise modeling

In section 3.2 identical additive process noise was introduced in the  $z$ -coordinates, as can be seen in (12).

The addition of identical process noise to each feature point in a rigid body dynamic system can be interpreted as the modeling of actuating forces or other influences affecting all of the observed points in a way that preserves the rigid structure. This implies however, that the process noise should be added identically to each point in the  $x$ -coordinates, i.e. as

$$\begin{aligned} dx_i &= (Ax_i + b) dt + Bd\beta_w \\ d\bar{y}_i &= \frac{1}{x_{i,3}} Cx_i dt + Dd\beta_{vi}, \end{aligned} \quad (23)$$

In this case, also the process noise should be transformed when applying the transformation (4).

In integrating stochastic differential equations one needs to make a choice on the interpretation of the integral of the noise term, which in turn has implications e.g. for transformations performed on the state variables. One possible interpretation of the integral is the Itô interpretation [16]. For a stochastic differential

$$dx(t) = f(t)dt + G(t, x(t))d\beta_w, \quad (24)$$

and a transformation of the type

$$z_j(t) = U_j(t, x(t)),$$

where  $U_j$  are scalar functions that are twice continuously differentiable in  $x$  and one time continuously differentiable in  $t$ , this choice of interpretation then leads to the Itô formula

$$dz_j = \left[ \frac{\partial U_j}{\partial t} + f^T \nabla U_j + \frac{1}{2} \text{tr}(GG^T \nabla[\nabla U_j]) \right] dt + \nabla U_j^T G d\beta_w. \quad (25)$$

Here  $\nabla U_j$  and  $\nabla[\nabla U_j]$  denotes the gradient of  $U_j$  and the matrix of second partial derivatives of  $U_j$  respectively, and  $\text{tr}$  means the trace of a matrix.

For the transformation (4) applied to the system (23)

we obtain the following components related to (25)

$$\frac{\partial U_j}{\partial t} = 0, \quad j = 1, 2, 3$$

$$\nabla U_1 = (z_{i,3} \quad 0 \quad -z_{i,1}z_{i,3})^T$$

$$\nabla U_2 = (0 \quad z_{i,3} \quad -z_{i,2}z_{i,3})^T$$

$$\nabla U_3 = (0 \quad 0 \quad -z_{i,3}^2)^T$$

$$\nabla[\nabla U_1] = \begin{pmatrix} 0 & 0 & -z_{i,3}^2 \\ 0 & 0 & 0 \\ -z_{i,3}^2 & 0 & 2z_{i,1}z_{i,3}^2 \end{pmatrix}$$

$$\nabla[\nabla U_2] = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -z_{i,3}^2 \\ 0 & -z_{i,3}^2 & 2z_{i,2}z_{i,3}^2 \end{pmatrix}$$

$$\nabla[\nabla U_3] = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2z_{i,3}^3 \end{pmatrix}.$$

Further, for simplicity assuming that  $B$  is diagonal with all elements equal to  $\sigma$ , and using the Itô formula (25), yields the transformed system

$$\begin{aligned} dz_i &= [\bar{A}z_i + \bar{b} + (\bar{c}^T z_i)z_i + \sigma^2 z_{i,3}^2 z_i] dt + \sigma S(z_i) d\beta_w \\ d\bar{y}_i &= Cz_i dt + Dd\beta_{vi}. \end{aligned} \quad (26)$$

with

$$S(z_i) = \begin{pmatrix} z_{i,3} & 0 & -z_{i,1}z_{i,3} \\ 0 & z_{i,3} & -z_{i,2}z_{i,3} \\ 0 & 0 & -z_{i,3}^2 \end{pmatrix}. \quad (27)$$

Clearly the transformation destroys the additive input noise structure assumed for the rigid body model. The analysis of filters for systems such as (26), especially in the context of connected filters and improved structure estimation, is a subject for future research.

## 6. Conclusions

Three-dimensional structure information can be estimated from two-dimensional images. In this paper, we have investigated the use of a nonlinear connected filter for recursive structure estimation in a rigid body motion perspective vision system, possibly affected by both measurement noise and process noise.

The filter is derived using the EKF algorithm for a transformed perspective dynamic system. A comparison between a connected filter and  $n$  independent filters, shown in Fig. 2 and Fig. 3, indicates that the estimation error

performance is improved when more feature points are included in the connected filter. The comparison was done using identical process noise for all feature points in the transformed system.

A theoretical analysis of an idealized linear system with identical process noise for all states is presented. Analytical expressions for the estimation error variance are derived. It can be seen how the estimation performance is affected by the noise properties and the number of points used in the connected filter.

An alternative model, which more accurately reflects the properties of a rigid body system, can be obtained by using identical process noise in the untransformed system. This requires however that the noise is included in the transformation. This can be done, using e.g. Itô's formula, leading to a stochastic system with nonlinear noise terms, for which it is not straightforward to use e.g. an extended Kalman filter.

Future work includes construction and investigation of the effect of connected filters on systems with more complex noise models. Another interesting problem concerns the influence of uncertainties in the motion parameters, and the relation to adaptive filtering techniques.

## References

- [1] Antonio Pedro Aguiar and João Pedro Hespanha. Minimum energy state estimation for systems with perspective outputs and state constraints. In *Proc. of CDC'03 - 42nd IEEE Conference on Decision and Control*, December 2003.
- [2] Ali Azarbayejani and Alex P. Pentland. Recursive estimation of motion, structure, and focal length. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(6):562–575, 1995.
- [3] Xinkai Chen and Hiroyuki Kano. A new state observer for perspective systems. *IEEE Transactions on Automatic Control*, 47(4):658–663, April 2002.
- [4] Ola Dahl, Fredrik Nyberg, Jan Holst, and Anders Heyden. Linear design of a nonlinear observer for perspective systems. In *Proc. of ICRA'05 - 2005 IEEE Conference on Robotics and Automation*, April 2005.
- [5] W. E. Dixon, Y. Fang, D. M. Dawson, and T. J. Flynn. Range identification for perspective vision systems. *IEEE Transactions on Automatic Control*, 48(12):2232–2238, December 2003.
- [6] 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.
- [7] Richard Hartley and Andrew Zisserman. *Multiple View Geometry*. Cambridge, 2003.
- [8] Andreas Huster. *Relative Position Sensing by Fusing Monocular Vision and Inertial Rate Sensors*. PhD dissertation, Stanford University, Department of Electrical Engineering, 2003.
- [9] Hiroshi Inaba, Akifumi Yoshida, Rixat Abdursul, and Bijoy K. Gosh. Observability of perspective dynamical systems. In *Proceedings of the 39th IEEE Conference on Decision and Control*, Sydney, Australia, December 2000.
- [10] Mrdjan Jankovic and Bijoy K. Ghosh. Visually guided ranging from observations of points, lines and curves via an identifier based nonlinear observer. *Systems & Control Letters*, 25:63–73, 1995.
- [11] Yi Ma, Stefano Soatto, Jana Košecá, and S. Shankar Sastry. *An Invitation to 3-D Vision*. Springer-Verlag, 2004.
- [12] L. Matthies, T. Kanade, and R. Szeliski. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3:209–236, 1989.
- [13] A. Matveev, X. Hu, R. Frezza, and H. Rehbinder. Observers for systems with implicit output. *IEEE Transactions on Automatic Control*, 45(1), January 2000.
- [14] P. S. Maybeck. *Stochastic Models, Estimation, and Control Volume 2*. Academic Press, 1982.
- [15] Richard M. Murray, Zexiang Li, and S. Shankar Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.
- [16] Bengt Öksendahl. *Stochastic Differential Equations*. Springer, 2000.
- [17] Henrik Rehbinder and Bijoy K. Gosh. Pose estimation using line-based dynamic vision and inertial sensors. *IEEE Transactions on Automatic Control*, 48(2), February 2003.
- [18] Stefano Soatto. 3-d structure from visual motion: Modeling, representation and observability. *Automatica*, 33(7):1287–1312, 1997.
- [19] Stefano Soatto, Ruggero Frezza, and Pietro Perona. Motion estimation via dynamic vision. *IEEE Transactions on Automatic Control*, 41(3), March 1996.