Dynamic structure from motion based on nonlinear adaptive observers

Ola Dahl and Anders Heyden
Applied Mathematics Group, Malmo University, Sweden
ola.dahl@mah.se, anders.heyden@mah.se

Abstract

Structure and motion estimation from long image sequences is a an important and difficult problem in computer vision. We propose a novel approach based on nonlinear and adaptive observers based on a dynamic model of the motion. The estimation of the three-dimensional position and velocity of the camera as well as the three-dimensional structure of the scene is done by observing states and parameters of a nonlinear dynamic system, containing a perspective transformation in the output equation, often referred to as a perspective dynamic system. An advantage of the proposed method is that it is filter-based, i.e. it provides an estimate of structure and motion at each time instance, which is then updated based on a novel image in the sequence. The observer demonstrates a trade-off compared to a more computer vision oriented approach, where no specific assumptions regarding the motion dynamics are required, but instead additional feature points are needed. Finally, the performance of the proposed method is shown in simulated experiments.

1. Introduction

One of the central problems in computer vision is the recovery of structure and motion from image sequences. Most approaches to this problem are batch-methods, where first all images are gathered and then the calculations are performed on all the data. These methods are usually based on multi-view tensors and nonlinear least squares optimization, see [8] for an overview of these approaches. However, some attempts have been made to develop recursive methods (in the sense of processing images as they becomes available and always having an estimate of motion and structure at hand), e.g. [16, 2] and also some related work in the area of automatic control, e.g. [6]. The main motivation for developing recursive methods is to being able to use them in real-time applications, where on-line structure and motion estimation is essential.

In order to apply nonlinear adaptive observers, a dynamic model of the motion of the camera is introduced. This formulation turns the structure and motion problem into a problem of observing states and parameters in the resulting dynamic perspective system. Structure and motion estimation without knowledge of motion parameters can be considered the most challenging case, and is described e.g. in [15], where structure-independent motion estimation is performed using a dynamic system, and in [14], where structure estimation is treated. References [15] and [14] present algorithms for estimating structure as well as motion using e.g. implicit extended Kalman filters. The algorithms are verified experimentally but it is difficult to establish analytical results regarding convergence and stability. A specific class of algorithms for structure estimation, where available values for angular and linear velocities are used and where position is estimated, can be formulated as *nonlinear observers*. This kind of observers are described e.g. in [13, 9, 3, 5, 4, 1, 10, 7], which present estimators for *structure only* using different kinds of nonlinear observers, providing various analytical results regarding stability, and simulation examples for illustration of observer performance.

This paper describes how a parametrization of the underlying dynamic system can be used to formulate estimation problems for *structure as well as motion*, and how the so obtained problem formulations can be used for the derivation of estimators, using available methods from nonlinear and adaptive control. Problem formulations for different estimation tasks are presented, and observers are derived and illustrated using simulation examples.

2. Dynamic perspective system

The dynamic system parametrization is derived from a dynamic system which is obtained, as is commonly done, from a description involving coordinate systems for the observed object and for the camera. For the purpose of clarity we also employ an inertial coordinate system when defining a specific motion.

The inertial coordinate system is denoted the a-system. The object coordinate system is denoted the b-system, and is attached to the observed object which is assumed to be a rigid body. The object may be stationary or moving. The camera coordinate system is referred to as the c-system, and is considered attached to a possibly moving camera.

A system of differential equations for the motion of the point p can be derived. Introducing the notation d_{aba} for the coordinates of the vector d_{ab}, corresponding to the centre of projection, when expressed using the orientation of the a-system, and the notation x_{bpb}, corresponding to a scene point, for the coordinates of the vector x_{bp} when expressed using the orientation of the b-system, we get, using a rotation matrix R_{ab} which expresses the relative orientation between the two coordinate systems, that $x_{apa} = d_{aba} + R_{ab}x_{bpb}$. Similarly, the coordinates of the vector x_{bp} can be expressed using the orientation of the c-system using a rotation matrix R_{cb}, as

$$x_{cpc} = d_{cbc} + R_{cb}x_{bpb} . \tag{1}$$

Define the skew-symmetric matrix $S(v)$ associated with a vector $v \in \mathbb{R}^3$, using a cross-product with an arbitrary vector u as $S(v)u = v \times u$. Introduce also the angular velocity vector ω_{cbc} and the matrix $S(\omega_{cbc}) = \dot{R}_{cb}R_{cb}^T$. Differentiating (1) with respect to time and using the orthogonality property $R_{cb}^T R_{cb} = \mathrm{I}$, together with the assumption of rigid body motion which implies $\dot{x}_{bpb} = 0$, then results in

$$\dot{x}_{cpc} = S(\omega_{cbc})x_{cpc} + \dot{d}_{cbc} - S(\omega_{cbc})d_{cbc} . \tag{2}$$

Observe that the relation (2) holds for an arbitrary point p. The camera model used here is a frontal pinhole imaging model [12] with an image plane parallel to the x_1-x_2-plane of the c-system, a focal length f, an optical center which coincides with the origin of the c-system, a camera transformation matrix $C \in \mathbb{R}^{2 \times 2}$ and an offset vector $\delta \in \mathbb{R}^{2 \times 1}$. Introducing the vectors $y = \begin{pmatrix} y_1 & y_2 \end{pmatrix}^T$ and $\xi = \begin{pmatrix} \xi_1 & \xi_2 \end{pmatrix}^T = \begin{pmatrix} \frac{x_{cpc,1}}{x_{cpc,3}} & \frac{x_{cpc,2}}{x_{cpc,3}} \end{pmatrix}^T$ and defining $C_f = f \cdot C$, this results in 2D image coordinates expressed in vector form as

$$y = C_f\xi + \delta . \tag{3}$$

For the purpose of deriving a dynamic system, for which observers can be constructed, introduce the simplified notation

$$x = x_{cpc}, \ d = d_{cbc}, \ \omega = \omega_{cbc}, \ \xi = \begin{pmatrix} \frac{x_1}{x_3} & \frac{x_2}{x_3} \end{pmatrix}^T . \tag{4}$$

Further, define the matrix A and the vector b as $A = S(\omega)$, $b = \dot{d} - Ad$. equation (2) with the output vector y given by the camera model (3) then results in the

system

$$\dot{x} = Ax + b, \quad y = C_f\xi + \delta . \tag{5}$$

This model can also be extended to describe the motion and observation of multiple points x^i on the same rigid object, where the model parameters A, b, C_f and δ, as a result of the rigid body assumption and the use of a single camera, are common to all the points x^i.

3. Dynamic vision parametrization

Given x from (5), introduce the scalar parameter γ and the vector z by

$$\gamma = \frac{1}{\sqrt{x^T x}}, \quad z = \gamma x . \tag{6}$$

It can be seen from (6) that z is the unit vector in the direction of the 3D position x, and also that γ is the inverted distance to the feature point under consideration, i.e. the 3D point with coordinates given by x.

Differentiating γ in (6) with respect to time using (5) and the fact that $x^T Ax = 0$ since A is skew-symmetric, gives

$$\dot{\gamma} = -\gamma^2 z^T b . \tag{7}$$

Combining (5) with the definition of z in (6) and using (7), we further have that $\dot{z} = Az + b\gamma - z(z^Tb)\gamma$. Observing that ξ, according to (4) and by the definition of z in (6), also can be expressed as $\xi = \begin{pmatrix} \frac{z_1}{z_3} & \frac{z_2}{z_3} \end{pmatrix}^T$, the dynamic system (5) can be formulated as

$$\dot{z} = Az + (I - zz^T)b\gamma, \quad y = C_f\xi + \delta . \tag{8}$$

Now assume that the camera is *calibrated*, i.e. that C_f and δ in (3) are known. Also assume that C_f is invertible. Since y is measured, these assumptions imply that ξ can be assumed known. By (6) and the definition of ξ in (4) the vector z can also be expressed as

$$z = \frac{1}{\sqrt{\xi_1^2 + \xi_2^2 + 1}} \begin{pmatrix} \xi_1 & \xi_2 & 1 \end{pmatrix}^T . \tag{9}$$

Thus, since ξ is assumed to be a known measurement signal also z can be assumed known. Combining (7) with the first equation in (8), and introducing $g_0(z) = I - zz^T$ a dynamic system can be formulated as

$$\dot{z} = Az + g_0(z)b\gamma, \quad \dot{\gamma} = -\gamma^2 z^T b . \tag{10}$$

Again, this dynamic system can be formulated for several points. Equation (6) together with (10) constitute the desired dynamic vision parametrization. It is referred to as a parametrization rather than e.g. a coordinate transformation, since the transformation from x to z is not invertible. Instead the vector z, which, assuming a calibrated camera, is measurable according to (9),

can be regarded as an alternative form of image coordinates. More specifically, the parametrization (6) can be interpreted as a projection onto a spherical image surface.

4. Structure and Motion Estimation

Introduce a measurable vector $\eta \in \mathbb{R}^N$ together with a vector of unknown parameters $\theta \in \mathbb{R}^M$. A dynamic system, where the vectors η and θ together constitute the state vector, will be used as the basis for different estimation problems. The dynamic system is written as

$$\dot{\eta} = \psi(\eta) + \phi^T(\eta)\theta, \quad \dot{\theta} = \mu(\eta, \theta) \qquad (11)$$

where ψ, ϕ and μ are vector-valued functions, determined from the particular estimation problem under consideration. The matrix ϕ will be denoted the *regressor* matrix.

Assuming for a moment a known angular velocity ω, the problem of estimating the linear velocity b as well as the structure parameters γ, can be formulated using (11), however augmented with an additional *scaling condition*. The need for the scaling condition can be seen from (10), where the linear velocity b only appears in the product γb. Consequently, γ and b cannot be distinguished without the use of additional constraints. Introducing $\beta = \gamma b$, and assuming that the vector b evolves according to a dynamic system $\gamma \dot{b} = \mu_\beta(\beta)$, the dynamic system (10) is reformulated, using the equation for $\dot{\gamma}$ in (10), as

$$\dot{z} = Az + g_0(z)\beta, \quad \dot{\beta} = -((z)^T\beta)\beta + \mu_\beta(\beta) . \quad (12)$$

Define normalized values of the parameters γ as $\alpha = \frac{\gamma}{\gamma_0}$, where γ_0 denotes γ for an arbitrary point. The problem of estimating the linear velocity b as well as the structure parameters γ can be formulated using two dynamic systems having the form (11). The first dynamic system is formulated for the purpose of estimating β_0 for the selected point above, and the second dynamic system is formulated for the purpose of estimating α for the other points. For this estimation task, the relation $\beta = \gamma b = \frac{\gamma}{\gamma_0}\gamma_0 b = \alpha_0\beta_0$ is used to reformulate (12) for the purpose of estimating α, given an estimate of β.

The parameters to be estimated, i.e. the quantities β_0 and α, need to be combined with a scaling condition, for the purpose of computing the structure parameters γ. A scaling condition can be derived e.g. from assumed knowledge of the distance between two object points. Assuming a distance d between two points x_0 and x_1, i.e. $(x_0 - x_1)^T(x_0 - x_1) = d^2$, we get, using $z = \gamma x$ according to (6)

$$(z_0 - \frac{1}{\alpha_1}z_1)^T(z_0 - \frac{1}{\alpha_1}z_1) = (\gamma_0)^2 d^2 \qquad (13)$$

Since z_0 and z_1 are measurable, equation (13) shows that γ_0 can be computed, given an estimated value of α_1 and an assumed value of the distance d. The remaining γ can then be computed.

In the case of unknown angular velocity the first dynamic system is extended. The parameter vector becomes $\theta_1 = \left(\omega^T \quad (\beta_0)^T\right)^T$ and the regressor matrix is given by $\phi_1(\eta_1) = \left(-S(z^1) \quad g_0(z^1)\right)^T$. The second dynamic system uses the estimated β_0 as well as the estimated ω which then replaces a known angular velocity with an estimated angular velocity.

5. Nonlinear and adaptive observers

Introduce a matrix F which is Hurwitz, and a symmetric positive definite matrix Q. A symmetric positive definite matrix P can then be computed as the unique solution to the Lyapunov equation [11],

$$F^T P + PF = -Q . \qquad (14)$$

Introducing the estimated quantities $\hat{\eta}$ and $\hat{\theta}$, an estimator for (11) can then be formulated as

$$\begin{aligned}\dot{\hat{\eta}} &= \psi(z) + F(\hat{\eta} - \eta) + \phi^T(\eta)\hat{\theta} \\ \dot{\hat{\theta}} &= -\phi(\eta)P(\hat{\eta} - \eta) + \mu(\eta, \hat{\theta}) .\end{aligned} \qquad (15)$$

The estimator (15) constitutes an extension of the estimator presented in [17]. The extension is here due to θ not being a constant parameter, as is assumed in [17]. Therefore, the second equation in (15) contains a correction term $\mu(\eta, \hat{\theta})$ which is not present in the estimator in [17].

The estimator (15) is formulated with reference to the dynamic system (11). Section 4 describes how the dynamic system (11) can be used as the basis for formulating the structure and motion estimation problem in dynamic vision.

The matrices F and Q can be regarded as *tuning parameters* for the estimator (15).

6. Simulation examples

The performance of the estimator (15) is demonstrated in simulation examples. In the simulations, we use $F = -10 \cdot I$ and $Q = 750 \cdot I$ and P is calculated from (14).

A *structure and motion estimation* example is presented, using the estimator (15) applied to the dynamic system (11), for estimation of angular and linear velocity as well as three-dimensional position. The observed object contains two feature points, executing a periodic motion used also in [3], governed by the parameter vectors $\omega = \left(-0.4 \quad 0.5 \quad 4\right)^T$
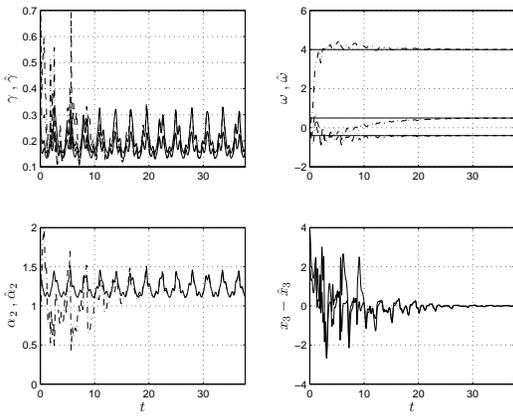
Figure 1. Structure and motion estimation results, for a periodic object motion as in [3]. True (solid) and estimated (dash-dotted) values of γ (upper left), α_1 (lower left) and ω (upper right). The depth errors are shown in the lower right plot.

and $b = \begin{pmatrix} 0 & 2\pi\sin(2\pi t) & 2\pi\cos(2\pi t) \end{pmatrix}^T$. The initial values for the object points are given by $x_0 = \begin{pmatrix} 2 & 2 & 4 & 2 & 2 & 2 \end{pmatrix}^T$, with a distance $d = 2$ between the two points. The estimation results are shown in Fig. 1, where it can be seen in the lower right plot that the three-dimensional position is recovered. It can also be seen, in the upper right plot, that the estimated angular velocity converges to its correct value. The upper left plot shows that the estimated parameters $\hat{\gamma}$ converge to the actual parameters γ for the two points, as is the case for α_1 as seen in the lower left plot.

7. Conclusions

Estimation of 3D structure and motion from 2D images can be performed using a dynamic systems formulation, where nonlinear and adaptive observers can be utilized for estimation of states and parameters. In this paper we have demonstrated how a single parametrization of the underlying perspective dynamic system can be used for formulation of estimation problems for structure as well as motion. The proposed nonlinear observer is able to estimate structure and motion, using as few as two feature points. This, however, requires that certain dynamic properties of the angular and linear velocities are known. The observer thus demonstrates a trade-off compared to a more computer vision oriented approach, where no specific assumptions regarding the motion dynamics are required, but instead additional feature points are needed.

References

[1] R. Abdursul, H. Inaba, and B. K. Ghosh. Nonlinear observers for perspective time-invariant linear systems. *Automatica*, 40:481–490, 2004.

[2] A. Azarbayejani and A. P. Pentland. Recursive estimation of motion, structure, and focal length. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(6):562–575, 1995.

[3] X. Chen and H. Kano. State observer for a class of nonlinear systems and its application to machine vision. *IEEE Transactions on Automatic Control*, 49(11):2085–2091, November 2004.

[4] O. Dahl, F. Nyberg, J. Holst, and A. 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] B. K. Ghosh, H. Inaba, and S. Takahashi. Identification of riccati dynamics under perspective and orthographic observations. *IEEE Transactions on Automatic Control*, 45(7):1267–1278, July 2000.

[7] S. Gupta, D. Aiken, G. Hu, and W. E. Dixon. Lyapunov-based range and motion identification for a nonaffine perspective dynamic system. In *American Control Conference*, June 2006.

[8] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University Press, 2003.

[9] M. Jankovic and B. 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.

[10] D. Karagiannis and A. Astolfi. A new solution to the problem of range identification in perspective vision systems. *IEEE Transactions on Automatic Control*, 50(12):2074–2077, December 2005.

[11] H. K. Khalil. *Nonlinear Systems*. Pearson Education, Inc., 2000.

[12] Y. Ma, S. Soatto, J. Košecká, and S. S. Sastry. *An Invitation to 3-D Vision*. Springer-Verlag, 2004.

[13] 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.

[14] S. Soatto. 3-d structure from visual motion: Modeling, representation and observability. *Automatica*, 33(7):1287–1312, 1997.

[15] S. Soatto, R. Frezza, and P. Perona. Motion estimation via dynamic vision. *IEEE Transactions on Automatic Control*, 41(3), March 1996.

[16] S. Soatto and P. Perona. Reducing structure from motion: A general framework for dynamic vision, part 1: Modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(9), September 1998.

[17] A. Teel, R. Kadiyala, P. Kokotovic, and S. Sastry. Indirect techniques for adaptive input-output linearization of non-linear systems. *Int. Journal of Control*, 53(1):192–222, 1991.