Recursive Estimation of Camera Motion from Uncalibrated Image Sequences

Stefano Soatto† Ruggero Frezza‡ Pietro Perona†‡
† California Institute of Technology 116-81, Pasadena–CA 91125
‡ Università di Padova, Dipartimento di Elettronica, Padova–Italy
{soatto,frezza,perona}@systems.caltech.edu

Technical Report **CIT-CDS 94-003**, California Institute of Technology
February 10, 1994

Abstract

In This memo we present an extension of the motion estimation scheme presented in a previous CDS technical report [14, 16], in order to deal with image sequences coming from an uncalibrated camera. The scheme is based on some results in epipolar geometry and invariant theory which can be found in [6]. Experiments are performed on noisy synthetic images.

1 Introduction

Camera motion estimation is a key task in many applications ranging from image compression, to autonomous vehicle navigation, to recognition. Motion estimation from image sequences is usually performed in two steps: **first the camera is calibrated**, in order to establish metric relationships between world coordinates and image-plane measurements. The internal parameters (pixel size, optical center, focal length), are usually estimated *off-line*. Once calibration is performed, we can **estimate camera motion** and ambient structure recursively from the image sequence in a variety of ways [1, 13, 17, 11].

Most of the recursive motion estimation schemes rely on exact knowledge of internal camera parameters. However, experimental evidence shows that these can change drastically during a long sequence [3] due to zooming and changing of the aperture. Moreover, *often it is not possible to access the physical device* which produced the sequence in order to calibrate it. A motion estimation scheme should therefore be able to estimate camera calibration while processing the sequence and estimating motion and structure.

Many approaches for camera calibration are available in the literature; they can roughly be classified as:

1. Batch schemes, *known structure* by including a calibration rig in the field of view (see [9] and references).

2. Active devices, *known motion* by controlling the configuration (pose) of the camera [4, 3, 2].

3. *Arbitrary structure and motion.* Camera self-calibration is performed along with motion estimation [6].

The first two approaches assume that the camera is available for measurements, by either controlling its motion or inserting a known object into the field of view. Therefore it seems that the third approach is the only feasible solution when the the device which produced the sequence is not available, as for example in image compression applications or automation of image processing tasks for the movie industry.

Faugeras et al. [6] propose a *batch* scheme which reconstructs the epipolar transformation of the camera, and then imposes the structure of such a transformation by solving a set of polynomial equations, known as Kruppa's equations [6]. However, the scheme has some substantial drawbacks which make it unattractive for real world applications. In particular

- High sensitivity to pixel-noise
- Numerical instability
- Motion parameters and internal parameters are treated alike. While camera-motion can vary arbitrarily during a sequence, it is conceivable that some parameters (for example the pixel size or aspect ratio) are constant over long periods of time
- Not all the information coming from a sequence is exploited. The scheme processes 3 images at a time and does not use temporal coherence (recursion) or a-priori information (such as range values for focal length, initial confidence in the position of the optical center etc.).

Hence we want a recursive scheme which, after each incoming image, updates the computation performed at the previous step. We also want the scheme to be *causal* so that it can be used for real-time implementations. Azarbayejani et al. [1] perform *partial calibration* by updating the focal length of the camera on-line together with camera motion.

To our knowledge, the problem of estimating camera motion and calibration recursively from an image sequence has never been addressed in the literature before.

In this paper we present a scheme for performing ego-motion estimation and camera calibration recursively and causally for an image sequence. It does not need a calibration rig nor to control motion, while it exploits redundancy at each step and computations from each previous step by recursion. A priori information about calibration can be used, if available. Internal parameter time constants are adjustable by tuning their random walk models.

The scheme is based on a recent method for recursive motion estimation [16], extended to estimate camera parameters according to the representation of [6]. A key feature of our scheme is that *the structure of the epipolar geometry is imposed explicitly as the structure of the state-space of the filter*, so we do not need to solve complicated polynomial equations in order to enforce such a structure. From a different point of view, our filter can be viewed as a recursive differential scheme for solving Kruppa's equations.

We report some experiments on noisy synthetic image sequences, and are in the process of testing the scheme on real image sequences. The results of the simulations are very promising in terms of accuracy, robustness and computational expenses.

2 Formulation of the scheme

2.1 Camera Model: internal parameters and ego-motion

The camera may be model as a perspective map $M : \mathbb{R}^3 \rightarrow \mathbb{R}^2$. The simplest instance is the so called "pinhole model": $\mathbf{X} \doteq \left[\begin{array}{ccc} X & Y & Z \end{array} \right]^T \mapsto \left[\begin{array}{cc} x & y \end{array} \right]^T \doteq \left[\begin{array}{cc} \frac{X}{Z} & \frac{Y}{Z} \end{array} \right]^T \doteq \mathbf{x}$. It can also be represented as a *linear map* between real projective spaces, $\bar{M} : \mathbb{R}P^3 \rightarrow \mathbb{R}P^2$: in homogeneous coordinates it is represented by a 3×4 matrix $\left[\begin{array}{c|c} \mathbf{A} & 0 \end{array} \right]$ where $\mathbf{A} \doteq \left[\begin{array}{ccc} fs_x & 0 & -i_0 \\ 0 & fs_y & -j_0 \\ 0 & 0 & 1 \end{array} \right]$ is the internal parameter matrix. f is the focal length, (i_0, j_0) the coordinates of the optical center and (sx, sy) the pixel sizes along the image plane coordinates. The deviation from 90^o of the angle between the optical axis and the ccd surface is usually on the order of 1^0, and we may therefore neglect it.

As the camera moves inside the (static) scene, the points move in its reference according to the rigid motion constraint: $\mathbf{X}(t+1) = R(t)\mathbf{X}(t) + T(t)$, where (R, T) represent the discrete camera motion. The goal of a self-calibrating motion scheme is to estimate the internal parameters and camera motion from the time-varying projection $\mathbf{x}(t)$ of a number of feature points.

2.2 The essential constraint and the fundamental matrix

Longuet-Higgins [10] introduced a simple coplanarity constraint which links the projective coordinates $\mathbf{x}$ of a point at time t, the corresponding $\mathbf{x}'$ at $t+1$, and the traslation T undergone by the camera:

$$\mathbf{x}'^T_i \mathbf{Q} \mathbf{x}_i = 0 \qquad \forall i = 1 \ldots N. \tag{1}$$

where $\mathbf{Q} \doteq RS \doteq R(T\wedge)$ is called the essential matrix. Given a number of such constraints, it is possible to estimate the motion which generated it [10, 18, 12, 5]. It can be proved easily that a 3×3 matrix is essential if and only if it has two equal singular values and zero determinant [12].

In the case of an uncalibrated camera, a similar constraint can be derived based on the epipolar geometry: given $\tilde{\mathbf{x}}(t)$ at time t, its correspondent at $t+1$, $\tilde{\mathbf{x}}(t+1)$, must lie on the epipolar line ${}^te_{t+1}$. Such a line is described in projective coordinates by a linear function of $\tilde{\mathbf{x}}(t)$. The representing matrix is called the *fundamental matrix*: ${}^te_{t+1} \doteq \mathbf{F}\tilde{\mathbf{x}}(t)$. It can be shown [6] that $\mathbf{F} \doteq \mathbf{A}^{-T}\mathbf{Q}\mathbf{A}^{-1}$, where $\mathbf{Q}$ is the essential matrix. From the definition of the epipolar line, we may derive a generalization of the essential constraint [6]:

$$\tilde{\mathbf{x}}'^T_i \mathbf{F} \tilde{\mathbf{x}}_i = 0 \qquad \forall i = 1 \ldots N. \tag{2}$$

The scheme presented in [6] consists in first estimating $\mathbf{F}$ from (2), and then impose its structure via solving the Kruppa equations, which correspond to enforce the fact that $\mathbf{A}^T\mathbf{F}\mathbf{A}$ (is essential and therefore) has two equal singular values and zero determinant.

2.3 The essential filter extended to fundamental matrices

The essential filter is a motion estimation paradigm recently presented in [16]. It solves motion estimation as identification of the exterior differential system determined by the essential constraint:

$$\begin{cases} \mathbf{x}_i^T(t+1)\mathbf{Q}(t)\mathbf{x}_i(t) = 0 \\ \tilde{\mathbf{x}}_i(t) = \mathbf{x}_i(t) + n_i(t) \end{cases} \qquad \forall i = 1 : N \tag{3}$$

We propose to extend the essential filter to estimate fundamental matrices, and *impose the structure of the fundamental matrix explicitly* by writing the estimator in local coordinates: the estimate at each step determines a matrix which is *fundamental by construction*, and we do not need to enforce the structure by solving poorly conditioned polynomial equations. The structure of resulting update is very similar to the essential filter; for details on the derivation see [15]:

$$\begin{bmatrix} \hat{\xi} \\ \hat{T} \\ \hat{R} \end{bmatrix}(t+1) = \begin{bmatrix} \hat{\xi} \\ \hat{T} \\ \hat{R} \end{bmatrix}(t) + L(t)\begin{bmatrix} \vdots \\ \tilde{\mathbf{x}}_i^T(t)\mathbf{A}^{-T}(\hat{\xi})\mathbf{Q}(\hat{T},\hat{R})\mathbf{A}^{-1}(\hat{\xi})\tilde{\mathbf{x}}_i(t-1) \\ \vdots \end{bmatrix} \tag{4}$$

where $\xi \doteq [fs_x, fs_y, i_0, j_0]^T$; L has the structure of the gain of an Implicit Extended Kalman Filter (IEKF) [8, 7, 16].

The scheme has a strong system-theoretical motivation, which we do not report here for reasons of space.

3 Experimental Assessment

We report one set of simulations on a noisy synthetic sequences. In figure 1 we show the estimate of translation and rotation parameters. In figure 2 we show the estimates of internal parameters. Noise on the image-plane was one tenth of a pixel. Convergence is reached in about 100 frames. Each iteration consists of about 100 Kflops: an implementation using Matlab (not optimized) runs at $.6Hz$ on a Sparc 10-20. We are currently experimenting on real image sequences and higher noise levels. More detailed experiments are reported in [15].

4 Conclusions

We have presented a scheme for estimating ego-motion and camera calibration from an image sequence. The scheme is based on an Implicit Extended Kalman Filter in the manifold of fundamental matrices. The update is written in local coordinates, so that at each step the estimated state is a fundamental matrix by construction, and we do not need to enforce the structure by solving complicated polynomial equations. Simulations are presented on noisy synthetic image sequences.

References

[1] A. Azarbayejani, B. Horowitz, and A. Pentland. Recursive estimation of structure and motion using relative orientation constraints. *Proc. CVPR*, New York, 1993.

[2] A. Basu. Active calibration: anternative strategy and analysis. *Trans. of the IEEE conf. CVPR*, 1993.

[3] J.L. Crowley, P. Bobet, and C. Schmidt. Maintaining stereo calibration by tracking image points. *Trans. of the IEEE conf. CVPR*, 1993.

[4] F. Du and M. Brady. Self-calibration of the intrinsic parameters of cameras for active vision systems. *Trans. of the IEEE conf. CVPR*, 1993.

[5] O. Faugeras. *Three dimensional vision, a geometric viewpoint.* MIT Press, 1993.

[6] O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera self-calibration: theory and experiments. *Proc. of the ECCV92, Vol. 588 of LNCS, Springer Verlag*, 1992.

[7] A.H. Jazwinski. *Stochastic Processes and Filtering Theory.* Academic Press, 1970.

[8] R.E. Kalman. A new approach to linear filtering and prediction problems. *Trans. of the ASME-Journal of basic engineering.*, 35-45, 1960.

[9] R.K. Lenz and R.Y. Tsai. Techniques for calibration of the image center and scale factor for high accuracy 3d vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, Sept. 1988.

[10] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[11] L. Matthies, R. Szelisky, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *Int. J. of computer vision*, 1989.

[12] S. Maybank. *Theory of reconstruction from image motion*, volume 28 of *Information Sciences*. Springer-Verlag, 1992.

[13] J. Oliensis and J. Inigo-Thomas. Recursive multi-frame structure from motion incorporating motion error. *Proc. DARPA Image Understanding Workshop*, 1992.

[14] S. Soatto, R. Frezza, and P. Perona. Recursive motion estimation on the essential manifold. *Technical Report CIT-CDS 93-021 and CIT-CNS 32/93, California Institute of Technology*, 1993.

[15] S. Soatto, R. Frezza, and P. Perona. Recursive estimation of camera motion from uncalibrated image sequences. *Technical Report CIT-CDS 94-003, California Institute of Technology*, 1994.

[16] S. Soatto, R. Frezza, and P. Perona. Motion estimation on the essential manifold. *Proc. of the ECCV 94 – To appear in "Lecture Notes in Computer Sciences", Springer Verlag*, May 1994.

[17] S. Soatto, P. Perona, R. Frezza, and G. Picci. Recursive motion and structure estimation with complete error characterization. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, pages 428–433, New York, June 1993.

[18] J. Weng, T. Huang, and N. Ahuja. Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(5):451–476, 1989.

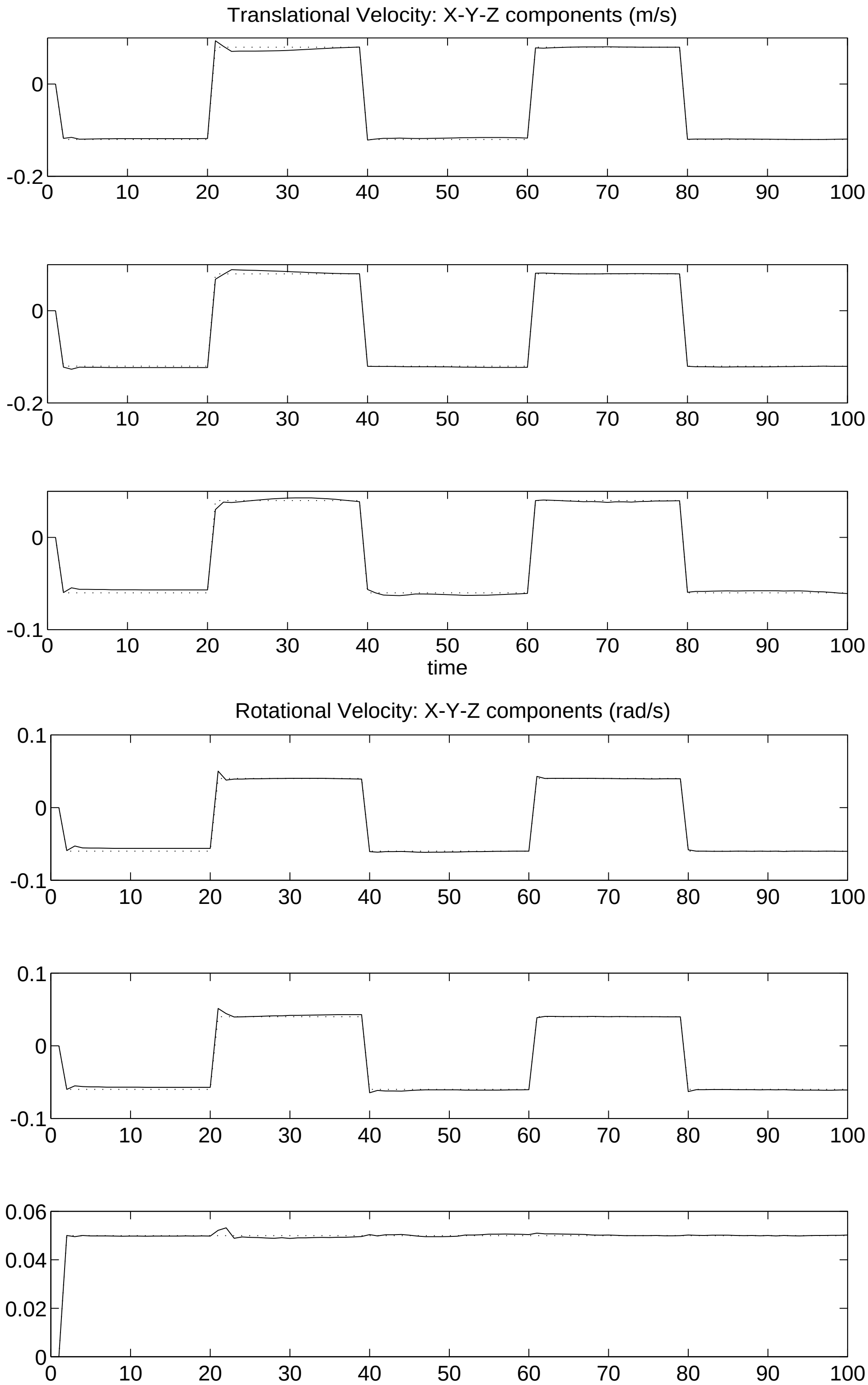

Figure 1: *(L) Translational velocity: filter estimates (solid) vs. true values (dotted) (R) Rotation*

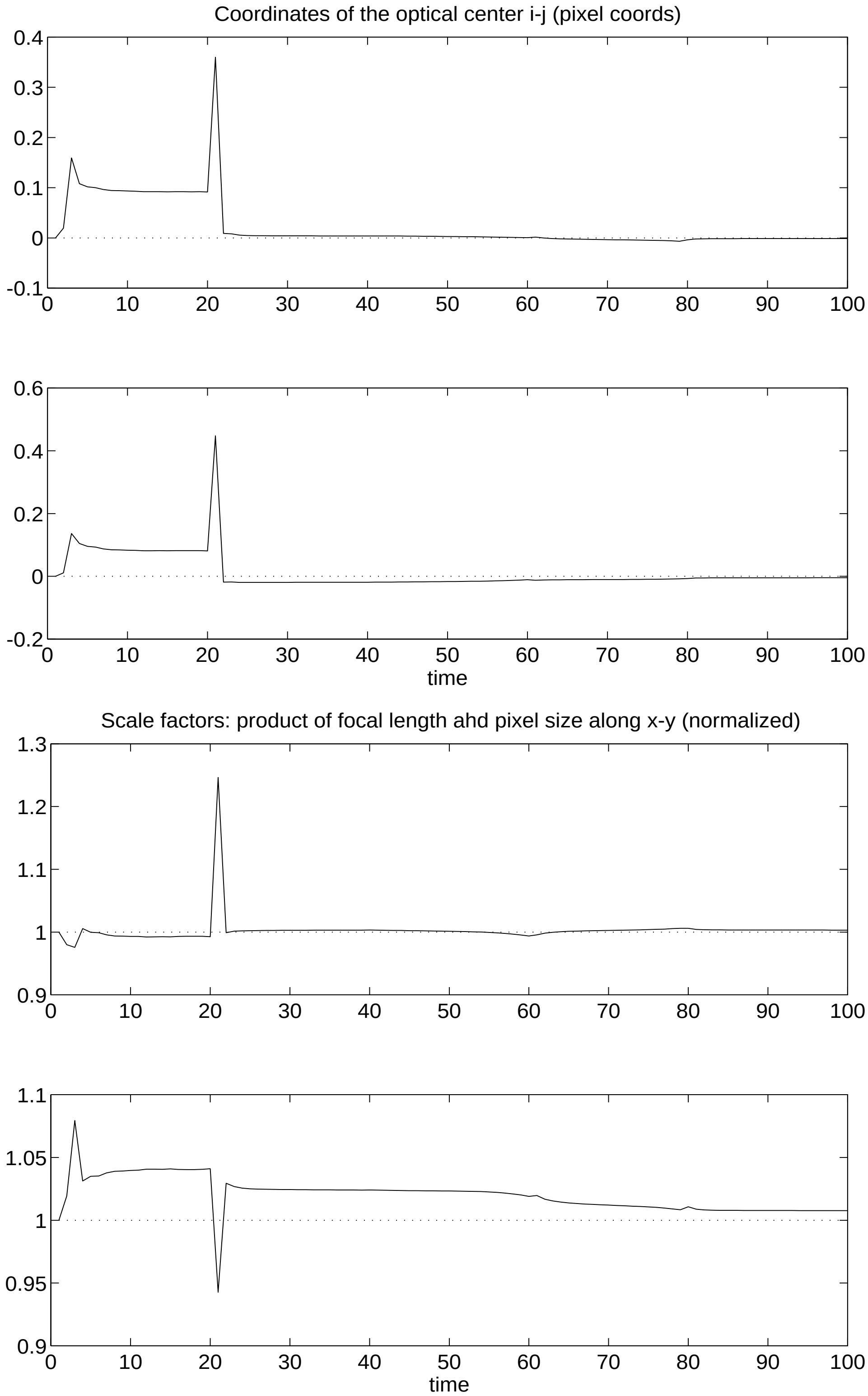

Figure 2: *(L) Coordinates of the center of projection: filter estimates vs. true values (R) Pixel size along image coordinates*