DYNAMIC VISUAL MOTION ESTIMATION FROM SUBSPACE CONSTRAINTS

Stefano Soatto† *and* *Pietro Perona*†‡

† California Institute of Technology 116-81, Pasadena–CA 91125 — soatto@caltech.edu
‡ Università di Padova, Dipartimento di Elettronica ed Informatica, Padova–Italy

ABSTRACT

The problem of estimating rigid motion from projections may be characterized using a nonlinear dynamical system, composed of the rigid motion constraint and the perspective map. The time derivative of the output of such a system, which is called the "motion field" and approximated by the "optical flow", is bilinear in the motion parameters, and may be used to specify a subspace constraint on either the direction of translation or the inverse depth of the observed points. Estimating motion may then be formulated as an optimization task constrained on such a subspace [4].

We pose the optimization problem in a system theoretic framework as the the identification of a nonlinear implicit dynamical system with parameters on a differentiable manifold, and use techniques which pertain to nonlinear estimation and identification theory to perform the optimization task in a principled manner.

The application of a general method presented in [12] results in a recursive and pseudo-optimal solution of the visual motion estimation problem, which has robustness properties far superior to other existing techniques we have implemented.

Experiments on real and synthetic image sequences show very promising results in terms of robustness, accuracy and computational efficiency.

1. MOTION ESTIMATION FROM A DYNAMIC MODEL

Let a scene be represented by a set of N feature points in 3D space moving rigidly with respect to the viewer; the "visual motion estimation" problem is *defined* by the rigidity constraint and the perspective projection equations. If $\mathbf{X}_i \doteq [X_i\ Y_i\ Z_i]^T$ are the coordinates of the i^{th} point and $\mathbf{x}_i \doteq [x_i\ y_i]^T$ the corresponding projections, we may write

$$\begin{cases} \dot{\mathbf{X}}_i = \Omega \wedge \mathbf{X}_i + V & \mathbf{X}(0) = \mathbf{X}_0 \\ \mathbf{x}_i = \pi(\mathbf{X}_i) + n_i & \forall i = 1:N \end{cases} \tag{1}$$

where n_i represents an error in measuring the position of the projection of the point i and π represents an ideal perspective projection. Solving the visual motion problem consists of estimating the ego-motion V, Ω from all the visible points, i.e. reconstructing the input of the above system from its measured output. We show that it is possible to invert the above system using a technique which has been recently introduced in [12] for identifying nonlinear implicit systems with parameters on a topological manifold.

The scheme is motivated by the work of Heeger and Jepson [4, 5] and may be considered as a recursive solution of their task using methods which pertain to the field of nonlinear estimation and identification theory. As a result, the minimization task which is the core of the subspace method for recovering rigid motion needs not to be performed by extensive search, as it is done in [4]. Instead, an Implicit Extended Kalman Filter (IEKF) [2, 7, 8, 12] is in charge of estimating the motion parameters recursively according to nonlinear prediction error criteria (for an introductory treatment of Prediction Error Methods (PEM) in a linear context, see for example [14]). As a result, our method exploits in a pseudo-optimal manner the information coming from a long stream of images, making the scheme robust and computationally efficient.

2. MOTION RECONSTRUCTION VIA INVERSION CONSTRAINED ON SUBSPACES

Consider the following expression of the first derivative of the output of the model (1), which is referred to as the "motion field":

$$\dot{\mathbf{x}}_i(t) = \left[\frac{1}{Z_i}\mathcal{A}_i \mid \mathcal{B}_i\right] \begin{bmatrix} V(t) \\ \Omega(t) \end{bmatrix} \tag{2}$$

where

$$\mathcal{A}_i \doteq \begin{bmatrix} 1 & 0 & -x_i \\ 0 & 1 & -y_i \end{bmatrix} \tag{3}$$

$$\mathcal{B}_i \doteq \begin{bmatrix} -x_i y_i & 1+x_i^2 & -y_i \\ -1-y_i^2 & x_i y_i & x_i \end{bmatrix}. \tag{4}$$

If we observe a sufficient number of points $\mathbf{x}_i\ \forall i = 1 \ldots N$, we can write an overdetermined system which can be solved for the inverse depth and the rotational velocity in a least-squares fashion. To this end, we rearrange the above equation as

$$\dot{\mathbf{x}}_i(t) = [\mathcal{A}_i V(\theta,\phi) \mid \mathcal{B}_i] \begin{bmatrix} \frac{1}{Z(t)_i} \\ \Omega(t) \end{bmatrix},$$

Research funded by the California Institute of Technology, a scholarship from the University of Padova, a fellowship from the "A. Gini" Foundation, an AT&T Foundation Special Purpose grant, ONR grant N0014-93-1-0990, grant ASI-RS-103 from the Italian Space Agency and the NSF National Young Investigator Award (P.P.). This work is registered as CDS Technical Report CIT-CDS 94-006, California Institute of Technology, January 1994 – revised February 1994.

where $V \in \mathbf{S}^2$ is represented in local (spherical) coordinates as $V(\theta, \phi)$. When we observe N points we can rearrange the above into a vector equality:

$$\dot{\mathbf{x}} = \tilde{C}(\theta, \phi)[\frac{1}{Z_1}, \ldots, \frac{1}{Z_N}, \Omega]^T, \tag{5}$$

where

$$\tilde{C}(\theta, \phi) \doteq \begin{bmatrix} \mathcal{A}_1 V & & \mathcal{B}_1 \\ & \ddots & \vdots \\ & & \mathcal{A}_N V \quad \mathcal{B}_N \end{bmatrix}$$

and $\mathbf{x}$ is a $2N$ column vector obtained by stacking the $\mathbf{x}_i \ \forall i = 1 \ldots N$ on top of each other. At this point we could solve the above equation in a least squares fashion for the inverse depth and the rotational velocity:

$$\begin{bmatrix} \frac{1}{Z_1} \\ \vdots \\ \frac{1}{Z_N} \\ \Omega \end{bmatrix} = \tilde{C}^{\dagger} \dot{\mathbf{x}}$$

where the symbol $\dagger$ denotes the pseudo-inverse. We can then plug the result into equation (5),

$$\dot{\mathbf{x}} = \tilde{C}(\theta, \phi) \tilde{C}^{\dagger} \dot{\mathbf{x}},$$

ending up with an *implicit constraint* on the direction of translation θ, ϕ. By rearranging the terms and writing explicitly the pseudo-inverse we get the following subspace algebraic constraint [4]:

$$\left[I - \tilde{C} \left(\tilde{C}^T \tilde{C} \right)^{-1}_{|\theta, \phi, \mathbf{x}} \tilde{C}^T \right] \dot{\mathbf{x}} \doteq \tilde{C}^{\perp} \dot{\mathbf{x}} = 0. \tag{6}$$

We can now try to approximate this constraint by solving the following nonlinear optimization problem:

$$\hat{V} = \arg \min_{V(\theta, \phi) \in \mathbf{S}^2} \| \tilde{C}^{\perp}(\theta, \phi) \dot{\mathbf{x}} \|. \tag{7}$$

In other words we are looking for the best vector in the sphere such that $\dot{\mathbf{x}}$ is the null space of the orthogonal complement of the range of $\tilde{C}$. If the matrix $\tilde{C}$ was invertible, the above constraint would be satisfied trivially for all directions of translation. However, when $2N > N + 1$, $\tilde{C}\tilde{C}^{\dagger}$ has rank at most $N + 1$, and therefore $\tilde{C}^{\perp}$ is not identically zero.

Note that we are trying to "adapt" the orthogonal complement of $\tilde{C}$, which is highly structured as a function of θ, ϕ, until a given vector $\dot{\mathbf{x}}$ is its null space. Heeger and Jepson [4] solve this problem by minimizing the two-norm of the above constraint using a extensive search over θ, ϕ, or a sampling of the sphere. This procedure does not exploit the geometric structure of the problem and is computationally expensive. Furthermore, it does not take into account the measurement noise, which enters into the minimization in a highly structured fashion. Temporal coherence of motion is also not taken into account: at each step we want to exploit all the processing performed at the previous time instant and update recursively the motion estimates.

In section 3 we rephrase the subspace constraints described in this section as a nonlinear and implicit dynamic model in exterior differential form [1]. Estimating motion corresponds to the identification of such a model with the parameters living on a sphere: we propose a principled solution for performing the optimization task. The method outputs motion estimates together with their reliability in the form of the second order statistics of the estimation error. Such an error may be used in subsequent modules for estimating structure [13].

2.1. Recovery of rotation and depth

Once the direction of translation has been estimated, we may compute the rotational velocity and inverse depth in a least-squares fashion from

$$\begin{bmatrix} \frac{1}{Z_1} \\ \vdots \\ \frac{1}{Z_N} \\ \Omega \end{bmatrix} = \tilde{C}^{\dagger}(\hat{\theta}, \hat{\phi}) \dot{\mathbf{x}}.$$

Note that, from the variance/covariance of the estimation error of the direction of translation θ, ϕ, we can characterize the second order statistics of the above "pseudo-measurement" of the rotational velocity. We may therefore design a simple linear Kalman filter based upon the model

$$\begin{cases} \Omega(t+1) = \Omega(t) + n_{rw}(t) \\ \tilde{C}^{\dagger}_{2N+1:2N+3}(\theta, \phi) \dot{\mathbf{x}} = \Omega(t) + n_{\Omega} \end{cases}$$

where n_{rw} is the noise driving the random walk model, which is to be intended as a tuning parameter, and n_{Ω} is an error whose variance can be easily inferred from the variance of θ, ϕ.

Once the rotational and translational velocity have been recovered, they may be fed, together with the variance of their estimation error, into a recursive structure from motion module which processes motion error, such as for example [10, 13].

2.2. Recovery of the mean distance

Note that the inverse depth of each point and the direction of translation play interchangeable roles, as it is evident from the structure of the motion field (2). We may therefore "pseudo-invert" the system (2) with respect to the direction of translation and the rotational velocity, and then perform a minimization similar to (7) with respect to the inverse depth of each point. Call $C_i \doteq \left[\frac{1}{Z_i} \mathcal{A}_i \mid \mathcal{B}_i \right]$, we have

$$\begin{bmatrix} \dot{\mathbf{x}}_1 \\ \dot{\mathbf{x}}_2 \\ \vdots \\ \dot{\mathbf{x}}_N \end{bmatrix} = \begin{bmatrix} \vdots \\ C_i \\ \vdots \end{bmatrix} \begin{bmatrix} V(t) \\ \Omega(t) \end{bmatrix} \Rightarrow \begin{bmatrix} V(t) \\ \Omega(t) \end{bmatrix} =$$

$$= \begin{bmatrix} \vdots \\ C_i \\ \vdots \end{bmatrix}^{\dagger} \begin{bmatrix} \dot{\mathbf{x}}_1 \\ \dot{\mathbf{x}}_2 \\ \vdots \\ \dot{\mathbf{x}}_N \end{bmatrix} \doteq C^{\dagger} \dot{\mathbf{x}}. \tag{8}$$

Note that C_i depends on the depth of the point Z_i, which we do not know. By substituting the above expression into eq. (2), we have an *implicit constraint* on Z_i [4], which we may approximate by solving w.r.t. Z_i the following optimization problem:

$$\hat{Z}_i = arg\min_{Z_i} \|\mathcal{C}^{\perp}\dot{\mathbf{x}}\|. \tag{9}$$

If $\mathcal{C}$ was invertible, again the above constraint would be satisfied trivially for all motions. However, when $2N > 3$, $\mathcal{C}\mathcal{C}^{\dagger}$ has rank at most three, and hence $(I - \mathcal{C}\mathcal{C}^{\dagger}) \neq 0$.

In many applications it is of interest to estimate the average distance of an object from the camera (depth of the centroid). For this case, it is sufficient to consider the minimization in eq. (9) when $Z_i = Z_c \ \forall i$; Z_c is the distance of the centroid.

3. SOLVING THE SUBSPACE OPTIMIZATION

Let us define $\alpha \doteq [\theta, \phi]^T$; $\mathbf{x}_i$ are measured up to some error, $\mathbf{y}_i \doteq \mathbf{x}_i + n_i \quad n_i \in \mathcal{N}(0, R_{n_i})$, which induces an error in the derivative: $\mathbf{y}_i' = \dot{\mathbf{x}}_i + n_i'$. Call $\mathbf{x}$ the column vector obtained by stacking the components of $\mathbf{x}_i$, similarly with $\dot{\mathbf{x}}$. Now define $\tilde{\mathcal{C}}^{\perp}(\mathbf{x}, \alpha) \doteq \left[I - \tilde{\mathcal{C}}\left(\tilde{\mathcal{C}}^T\tilde{\mathcal{C}}\right)^{-1}_{|\alpha,\mathbf{x}}\tilde{\mathcal{C}}^T\right]$. Then the subspace constraint (6) may be written as $\tilde{\mathcal{C}}^{\perp}(\mathbf{x}, \alpha)\dot{\mathbf{x}} = 0$. Now

$$\begin{cases} \tilde{\mathcal{C}}^{\perp}(\mathbf{x}, \alpha)\dot{\mathbf{x}} = 0 & V(\alpha) \in \mathbf{S}^2 \\ \mathbf{y}_i \doteq \mathbf{x}_i + n_i & \forall\, i = 1 \ldots N \end{cases}$$

represents a nonlinear implicit system of a particular class, called Exterior Differential Systems [1]. *Solving for translation is equivalent to identifying the above exterior differential system with parameters on a differentiable manifold* (the sphere in this case) from the noisy data $\mathbf{y}$.

We have addressed this problem using a general framework presented in [12]. The solution is given by the simple iteration

Prediction step

$$\begin{cases} \hat{\alpha}(t+1|t) = \hat{\alpha}(t|t) & \hat{\alpha}(0|0) = \alpha_0 \\ P(t+1|t) = P(t|t) + R_{\alpha}(t) & P(0|0) = P_0 \end{cases}$$

Update step

$$\begin{cases} \hat{\alpha}(t+1|t+1) = \hat{\alpha}(t+1|t) + \\ \quad L(t+1)\tilde{\mathcal{C}}^{\perp}(\mathbf{y}(t), \alpha(t+1|t))\mathbf{y}' \\ P(t+1|t+1) = \\ \quad \Gamma(t+1)P(t+1|t)\Gamma^T(t+1) + \\ \quad L(t+1)D_+(t)R_n(t+1)D_+^T(t)L^T(t+1) \end{cases}$$

where

$$\begin{cases} L(t+1) = P(t+1|t)C^T(t+1)\Lambda^{\dagger}(t+1) \\ \Lambda(t+1) = C(t+1)P(t+1|t)C^T(t+1) + \\ \qquad + D_+(t+1)R_n(t+1)D_+^T(t+1) \\ \Gamma(t+1) = I - L(t+1)C(t+1) \\ D_+(t+1) \doteq \left(\frac{\partial\tilde{\mathcal{C}}^{\perp}\dot{\mathbf{x}}}{\partial\mathbf{x}(t)}\right)_{|\mathbf{y}(t),\hat{\alpha}(t)} \\ C(t+1) \doteq \left(\frac{\partial\tilde{\mathcal{C}}^{\perp}\dot{\mathbf{x}}}{\partial\alpha(t)}\right)_{|\mathbf{y}(t),\hat{\alpha}(t)} \end{cases}$$

the interested reader may find the detailed derivation in [12].

In order to be able to assess the convergence of the above scheme, we must prove its observability/identifiability. When translated into the language of dynamic estimation, the analysis of Heeger and Jepson [6] can be intended as the observability analysis of our method; in particular it shows that the scheme converges under general position conditions.

4. EXPERIMENTAL ASSESSMENT

We have experimented with the scheme on real and noisy synthetic image sequences. Instead of computing the pseudo-inverse of the variance of the innovation Λ, we have extended its rank by adding a small random matrix, which takes into account the linearization error and prevents the filter from saturating.

For the same data set used in [13], the scheme proves far more robust to the effect of measurement noise. Convergence is reached from *arbitrary initial condition* and noise in the image plane coordinates up to 8 pixel std for a field of view of approximately 40° (see figure 1). The scheme converges also with higher noise levels if properly initialized.

An estimate for more usual noise levels (one pixel std) is reported in figure 2. The least-squares estimates of rotational velocity are plotted in figure 3 (dashed lines), and compared with the recursive estimates (solid line).

The behavior of the filter in the presence of local minima is described in the technical report CIT-CDS 94-006[1].

4.1. Experiments with real image sequences

We have tested the scheme on real image sequences: the noise level achieved by the most common feature tracking/optical flow techniques is easily handled by the filter. As an example we report here the filter estimates for the rocket scene, for comparison with [11]. Due to the fact that the filter takes about 20 frames to converge, we have doubled the sequence and used one run as initial condition for the second run, which is displayed in image 6.

5. CONCLUSIONS

We have formulated a new recursive scheme for estimating rigid motion under perspective via identifying a nonlinear implicit dynamic model with parameters on a manifold. The motivation comes from the work of Heeger and Jepson [4], who propose to view motion estimation as an optimization problem constrained to a subspace.

Using results from nonlinear estimation and identification theory, we formulate a motion estimator which is fast, computationally efficient, accurate and more robust than any recursive motion estimation scheme we have implemented. Extensive experiments have been performed that highlight such features.

[1]This paper can be obtained via the Worldwide Net Mosaic (http://avalon.caltech.edu/cds/techreports/).

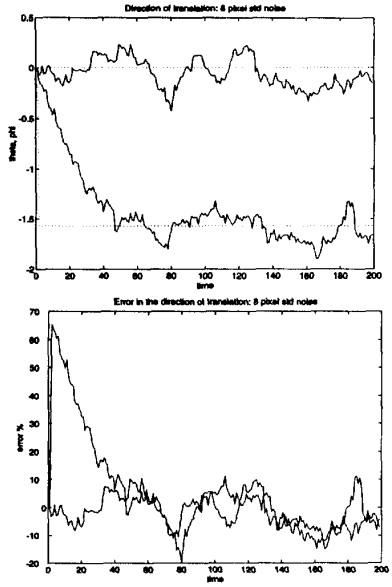

Figure 1: *(Top) Estimates of the two components of the direction of translation. The noise in the image plane measurements had 8 pixel standard deviation. The initial conditions were zero for both components. The ground truth is in dotted lines (Bottom) Estimation error for the direction of translation. With noise of 8 pixel std in the data, the estimates are still within 10 % of the true value.*

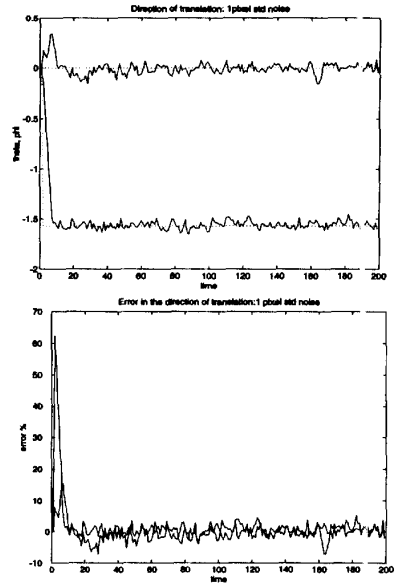

Figure 2: *Estimates and errors for the direction of translation when the noise in the image plane has a standard deviation of 1 pixel. Note that convergence is reached from zero initial condition in about 10 steps.*

6. REFERENCES

[1] Bryant, Chern, Goldberg, and Goldsmith. *Exterior Differential Systems.* Mathematical Research Institute. Springer Verlag, 1992.

[2] R.S. Bucy. Non-linear filtering theory. *IEEE Trans. A.C. AC-10, 198*, 1965.

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

[4] D. Heeger and A. Jepson. Subspace methods for recovering rigid motion i: algorithm and implementation. *Int. J. Comp. Vision vol. 7 (2)*, 1992.

[5] D. Heeger and A. Jepson. Subspace methods for recovering rigid motion i: algorithm and implementation. RBCV TR-90-35, University of Toronto – CS dept., November 1990. Revised July 1991.

[6] D. Heeger and A. Jepson. Subspace methods for recovering rigid motion ii: theory. RBCV TR-90-35, University of Toronto – CS dept., November 1990. Revised July 1991.

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

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

[11] S. Soatto, R. Frezza, and P. Perona. Motion estimation on the essential manifold. *In "Computer Vision, ECCV 94, Lecture Notes in Computer Sciences Vol. 801", Springer Verlag*, May 1994.

[12] S. Soatto, R. Frezza, P. Perona, and G. Picci. Motion estimation via dynamic vision. *Technical Report CIT-CDS 94-004 , California Institute of Technology. Reduced version to appear in Proc. of the 33rd IEEE conf. on Decision and Control. Submitted to the IEEE transactions on Automatic Control*, Feb. 1994.

[13] 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.

[14] T. Soderstorm and P. Stoica. *System Identification.* Prentice Hall, 1989.

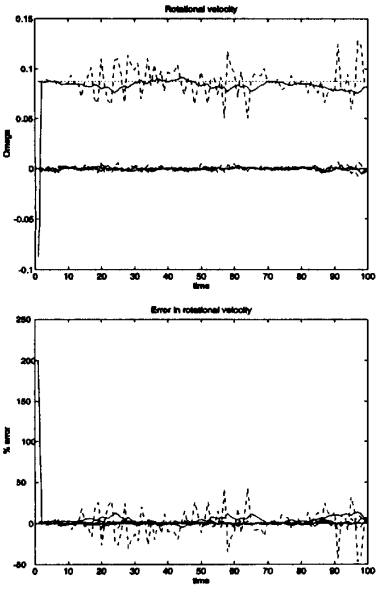

Figure 3: *Estimates for the components of rotational velocity (top) and corresponding error (bottom). Ground truth is displayed in dotted lines; the filtered estimates are in solid lines. The least-squares computation of the rotational velocity is in dashed lines.*

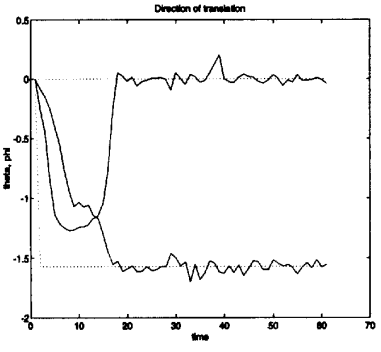

Figure 4: *Convergence to a shallow local minimum and then to the correct rigid motion.*

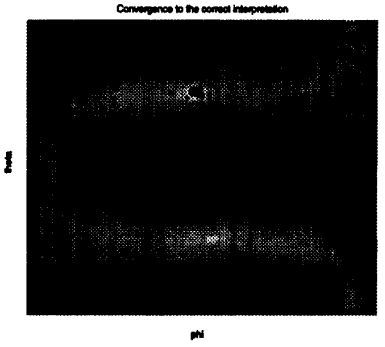

Figure 5: *Trajectory of the filter is superimposed to the average residual function (darker tones for larger residual, see figure 4).*

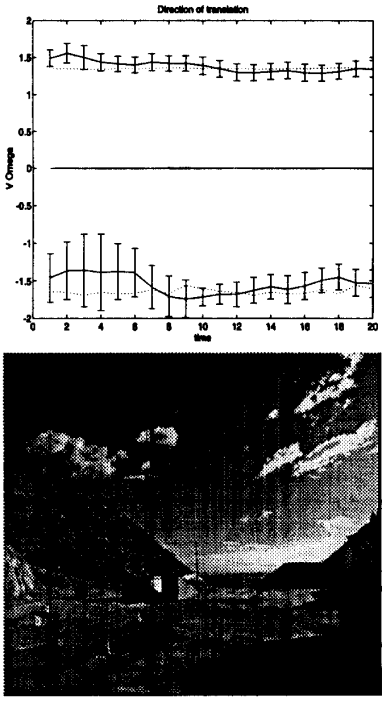

Figure 6: *(Top) Estimate of the direction of translation for the rocket scene. (Bottom) One image of the rocket scene.*