

Recursive Motion and Structure Estimation with Complete Error Characterization

S. Soatto† P. Perona†† R. Frezza† G. Picci† *

† California Institute of Technology 116-81 — Pasadena, CA 91125 — U.S.A. — soatto@caltech.edu

†† Università di Padova — Dipartimento di Elettronica ed Informatica — 35100 Padova — Italy

Abstract

We present an algorithm that performs recursive estimation of ego-motion and ambient structure from a stream of monocular perspective images of a number of feature points. The algorithm is based on an Extended Kalman Filter (EKF) that integrates over time the instantaneous motion and structure measurements computed by a 2-perspective-views step.

Key features of our filter are (1) global observability of the model, (2) complete on-line characterization of the uncertainty of the measurements provided by the two-views step. The filter is thus guaranteed to be well-behaved regardless of the particular motion undergone by the observer. Regions of motion space that do not allow recovery of structure (e.g. pure rotation) may be crossed while maintaining good estimates of structure and motion; whenever reliable measurements are available they are exploited. The algorithm works well for arbitrary motions with minimal smoothness assumptions and no ad hoc tuning. Simulations are presented that illustrate these characteristics.

1 Introduction

We consider the problem of reconstructing the geometrical structure of a rigid, static scene and the motion of the observer from a sequence of noisy monocular images representing perspective projections of the scene during motion. It is well known that two perspective projections onto a plane of five or more points in general configuration suffice to determine their position in space and the relative motion of the two projecting planes up to a scale factor [6, 16, 25].

When more points are observed one may exploit redundancy to find a solution that rejects noise minimizing the discrepancy from the rigidity assumption. Moreover it is possible to exploit the fact that the scene is static and integrate over time the information available from a long sequence of images. Many attempts have been made over the last decade to recover structure and motion using stochastic estimators with different degrees of generality and computational feasibility (see the introduction of [8] for a review), but to our judgement none of them is complete in terms of

*Research sponsored by NSF grant n. 51115, an AT & T Foundation technical special purpose grant, a donation by Dr. Tongsoo Park, and by ASI grant n. 92-RS-103

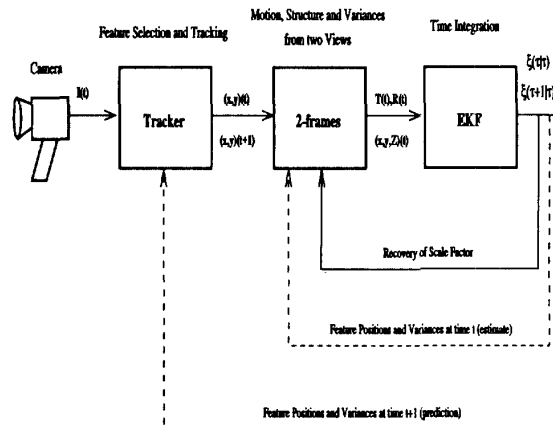


Figure 1: Structure of the algorithm.

conditioning with respect to motion and error treatment, and robust enough to be used in real-world applications.

We propose a recursive algorithm for calculating structure and motion whose state equation is globally observable so that the filter is well-conditioned regardless of the particular motion. Each component of the state is measured by a two-frames module together with the variance/covariance matrix of the measurement errors so that we need not use heuristic techniques or approximative information about the sensors' performance to set them.

Our algorithm is composed of three modules: the first module computes optical flow/features displacement at each new image frame, the second module computes the instantaneous structure and motion, the third module integrates in time the instantaneous measurements.

As a first module one may use any one of a number of optical flow/feature displacement algorithms described in the literature (see [2] for a survey with tests).

The second module calculates structure and motion in closed-form. We have chosen an algorithm recently presented by Hartley [9] for its simplicity and robustness. It is based on a technique proposed by Longuet-Higgins [16].

The third module is based on an Extended Kalman Filter (EKF) working at the same frequency as that of the image sequence.

Since the EKF requires specification of the variance/covariance matrix for the measurements, a key requirement of the second module is that, along with the instantaneous motion and structure measurements, good estimates of the *reliability* of these measurements are provided. A complete characterization of the error sensitivity of the second module is calculated in [18].

The overall structure of the algorithm is depicted in fig. 1. The inner loop is used to calculate the scale factor for the translation, and will be explained later.

2 Description of the algorithm

2.1 Rigid motion and notation

Let $\mathbf{X}^i = [X \ Y \ Z]^T$ be the coordinates of a generic point of a *static* scene with respect to an orthonormal reference frame centered in the pupil of the monocular observer. The Z axis points forward and coincides with the optical axis, while X and Y are parallel to the image plane, arranged in order to have a right-handed reference frame. As the observer moves in the scene between two time instants, the generic point moves in the observer's reference and its coordinates change according to

$$\mathbf{X}^i(t+1) = T(t) + R(t)\mathbf{X}^i(t) \quad (1)$$

where $T(t)$ is the position of the origin of the observer's reference at time t with respect to the reference at time $t+1$, while $R(t)$ is a rotation matrix whose columns are the coordinates of the reference axes at time t with respect to the same axes at $t+1$. Of course we could use different representations for orientation, as for example unit quaternions. Suppose we observe the points through their ideal perspective projection, assuming for simplicity unitary focal length. Then the coordinates of the projections are:

$$\mathbf{x}^i(t) \doteq \begin{bmatrix} x \\ y \end{bmatrix}^i(t) = \frac{\begin{bmatrix} X(t) \\ Y(t) \end{bmatrix}^i}{Z^i(t)} \quad \forall i = 1 : N, \quad (2)$$

which change according to (1) as the observer moves.

2.2 Recovery of structure as a filtering problem

Notice that in (1) we may interpret the quantity of interest as the state of a linear dynamical system, provided that we know the motion T, R exactly. In this case, if we measure the projection of the N observed feature points onto the image plane, $\mathbf{x}^i \forall i = 1 : N$, up to a certain error, we can then set up a measurement equation for the model and try to estimate its state from such measurements:

$$\mathbf{X}^i(t+1) = T(t) + R(t)\mathbf{X}^i(t) + m^i(t) \quad (3)$$

$$\tilde{\mathbf{x}}^i(t) = \mathbf{x}^i(t) + n^i(t) \quad (4)$$

where $i = 1 : N$, $n^i(t)$ is the measurement error. Observe that, although equation (1) is an exact representation for

rigid-stationary scenes, we have added here a noise term $m^i(t)$ which accounts for uncertainty in the knowledge of motion and round-off errors in the numerics.

If this system was linear and if the noises were ideal zero-mean, white, Gaussian distributed, then the Kalman Filter would be the optimal state estimator in the sense of minimum variance error [15]. A suboptimal version of the traditional Kalman Filter extended to non-linear systems, known as "Extended Kalman Filter (EKF)", has become very popular lately for its simplicity. Standard references on Kalman Filtering are [1, 14, 15]. A customary way to approach the problem of model errors is to assume zero-mean and whiteness, and play with variances of such errors until the prediction error is as uncorrelated as possible. This procedure is called "tuning", and standard tests can be used for the purpose.

2.3 Formulation of a filter for estimating structure: observability of the model

The filter described in the previous section may be reformulated so that the measurement equation is linear and the non-linearity is transferred to the state model. This is a more desirable situation, since the EKF treats the measurements as coming from a linear instrument. Furthermore the (small) linearization error introduced in the state equation contributes to keep the state variance non-zero and hence helps preventing saturation.

Suppose again, for the moment, known motion and rewrite the dynamical system describing the movement of the projected point p^i including in the state the unknown depth: call $\xi^i(t) \doteq [x \ y \ Z]^T(t)$, and $C \doteq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, then

$$\xi^i(t+1) = \mathcal{F}(\xi^i(t), T(t), R(t)) + m^i(t) \quad (5)$$

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix}^i(t) = C\xi^i(t) + n^i(t) \quad (6)$$

$\forall i = 1 : N$, where \mathcal{F} is given by

$$\mathcal{F}(\xi, T, R) \doteq \begin{bmatrix} C(T+R\xi) \\ \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T(T+R\xi) \\ \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T(T+R\xi) \end{bmatrix} \quad (7)$$

and $\mathbf{X} = [\xi_1 \ \xi_2 \ 1]^T \xi_3$. Then one may set up an EKF for the system described by the previous equations with state $\xi^T = [\xi_1^T, \dots, \xi_N^T, \dots]$.

This structure has been used, although in a different formulation, by Matthies et al. [17] and Heel [11, 12, 13]. It is important to notice that this approach is not general and presents some difficulties. First of all *motion is considered known*, second *the behavior of the filter is affected by the motion itself*: the observability of the $3i^{TH}$ component of the state (depth) in fact is conditioned by the motion, e.g. for small translations and distant points the filter tends to saturate selectively along the $3i^{TH}$ components.

2.4 Modified model for estimating structure and motion

The obvious way to get rid of observability problems is to

measure directly all of the components of the state. We need to make two modifications to the scheme just described: (1) insert the motion parameters in the state, (2) measure directly depth along with motion; these measurements may be made visually with a two-perspective-views algorithm (we use [9] in our experiments). In this case the observability matrix is the identity, and the filter is globally observable.

In order to include motion in the state of the filter we need to write a dynamical equation describing it: if the camera is mounted on a moving vehicle we may know the dynamical model for motion, otherwise we can use a statistical model, the simplest being a random walk. In the present paper we will use a first order random walk, which correspond to a mild smoothness hypothesis; indeed the model proves quite flexible and suitable to model also non-regular motion. The complete state and measurement equations are therefore:

$$\begin{bmatrix} \vdots \\ \xi^i \\ \vdots \\ T \\ \Omega \end{bmatrix} (t+1) = \begin{bmatrix} \vdots \\ \mathcal{F}(\xi^i, T, \Omega) \\ \vdots \\ T \\ \Omega \end{bmatrix} + m(t) \quad (8)$$

$$\begin{bmatrix} \vdots \\ \xi^i \\ \vdots \\ \tilde{T} \\ \tilde{\Omega} \end{bmatrix} (t) = \begin{bmatrix} \vdots \\ \xi^i \\ \vdots \\ T \\ \Omega \end{bmatrix} + n(t). \quad (9)$$

Notice that we have represented rotation via the vector Ω , instead of the matrix R . The two quantities are related by $R(\cdot) = e^{(\Omega \wedge)}(\cdot)$, where the operator $(\Omega \wedge)$ is a skew symmetric matrix which represents the wedge product $\Omega \wedge (\cdot)$. Notice that representing orientation via unit quaternions would avoid the transformation between R and Ω , but also hide the simple structure of the state equation.

Now the state equation is completely observable from the measurements, though of course we have not solved the problem of determining depth accurately in absence of reliable data, since the same uncertainty that used to affect the observability of the filter now affects the quality of the measurements in Z : intuitively if the (monocular) observer stops he is no longer able to resolve depth, and he feeds the filter with meaningless measurements.

In other words, we have traded observability for measurements quality. Notice though that in the Kalman Filter formalism we may model explicitly the accuracy of the measurements, so that this intrinsic difficulty may be handled in a principled way. The filter weights the measurements through their inverse variance and hence is able to discard unreliable data.

2.5 Two-view algorithm to measure depth and motion; recovery of the scale factor

In order to be able to use the above formulation we need a motion-and-structure algorithm that at each step calculates from visual data an instantaneous estimate of the depth of

each point and of the motion of the observer *together with the error variance* of those estimates.

In our experiments we use a scheme described in [18] that consists in a two-perspective views algorithm that, from the position in the image plane of $N \geq 8$ points, returns an estimate for the depth of such points and for the motion of the observer, and the variance of such estimates.

The 2-views algorithm consists of a chain of three Singular Value Decompositions (SVD) along the lines of [9, 16, 19, 23]. We give here a brief summary (we use Longuet-Higgins' notation – do not confuse the essential matrix Q with the variance of the model error Q and the rotation matrix R with R , the variance of the measurement errors):

1. A set of N equations $x_i^T Q x_i = 0$ is solved simultaneously via SVD to obtain the 'essential' matrix Q .
2. The matrix Q is factored into its rotation and translation components R and T using a second SVD. Ω is calculated from R using Rodrigues' formula.
3. Z_i , the depth of each point, is calculated using a Total Least Squares procedure implemented via a third SVD.

The error estimates are computed in the same spirit as Weng et al. [26]: the joint covariance matrix of the measurements T , Ω , and the Z_i , is computed by linearizing the algorithm around the zero-noise solution. This may be done by applying three times the equations coming from the linearization of the SVD [18]. The measurement error covariance matrix is computed at each iteration along with structure and motion and fed to the EKF.

Translation and depth can be resolved only up to a scale factor, due to the perspective projection framework. In order to overcome this ambiguity we need some external information such as the norm of the translation, or the distance between any two points in space, or the true depth of a point. Once this information is available *at one time instant* it can be propagated to the subsequent ones, exploiting the temporal coherence of the structure (see [21] for details).

3 Experimental assessment

3.1 Implementation and tuning of the filter

We have implemented the filter using `Matlab` and run computer simulations for testing it over a variety of motions and configuration of points. The equations of the filter are summarized in [14] (theorem 8.1 on page 278: substitute ξ for x , \mathcal{F} for f , C for $h[\cdot]$).

As initial condition we have chosen the first output of the 2-frames step, however the filter proved very insensitive, so that any other choice will also work (including zero), the only difference being a more appreciable transient. The variance of the measurement errors (R in [14]) are calculated in the two-frame step which is described in [18]. The only parameters which remain to be assigned are the components of the matrix variance of the model error (Q in [14]), for which the tuning procedure is carried out. In order to perform a coarse tuning we postulate, as customary, a diagonal Q and run the "Cumulative Periodogram" test with parameters varying by

one order of magnitude at a time. The filter proves fairly insensitive to tuning: we observe experimental convergence over three order of magnitude variations in the tuning parameters (though, of course, the quality of the estimates does vary). We would like to emphasize a consequence calculating measurement reliability at each time-step: *the tuning is independent of kind of motion* and has to be performed only once. The computational cost of the whole procedure, which consists of additions and multiplications of matrices, one linearization and a matrix inversion, is $O(3N + 7)^3$, where N is the number of observed points. For a detailed discussion of the implementation the reader is referred to [21].

3.2 A challenging motion

We have run a computer simulation for testing the algorithm on a variety of motions. In this section we present the simulations obtained for a non-smooth motion with singularities.

In our experiments we have considered a visual field of approximately 40° , and we have assumed standard deviation of the error in the localization of the features $n_0 = 2 \times 10^{-3}m$, which corresponds to 1 pixel, according to the performance of current feature tracking algorithms (see [2]). We have considered a set of $N=20$ points generated with a uniform random generator in a cube of side 1m with centroid 1.5m ahead of the observer, corresponding to a visual field of approximately 40° . The motion is explained in detail in the caption of figure 2. Note that motion has discontinuities which challenge the dynamical model we have used, a first order random walk, which represents some smoothness hypothesis. Indeed, as it can be seen from figures 2 and 3, the model is flexible, as it performs also in the case of non regular motion. Observe now that as the observer stops translating we have no parallax and the two-frame algorithm is no longer able to perceive depth. It is however able to perceive rotation. Notice that the impossibility of estimating depth in this case is inherent in the geometry, but that the filter needs not saturate: the algorithm retains its previous estimate of structure (depth) and updates them according to the (purely rotational) motion (see fig. 4). Notice that we do not use external information about the motion. Only the true distance between two features at time 0 is used to calculate the scale factor. Also note that any useful information about motion can be easily integrated within this framework, for example by inserting a state model for motion when available, or simply by including measurements for example from board instrumentation.

4 Relation to previous work

The literature on Kalman Filter-based recursive algorithms for recovering 3-D structure and/or motion from a monocular sequence of images dates back to the early eighties [7]. Extensive work has been done on three main variants of the problem: recovery of motion with known structure [4, 7, 8], recovery of structure with known motion [17, 20], and simultaneous recovery of motion and structure [3, 12, 13]. We review the literature related to our work in the next paragraph. We do not review here the extensive literature on

batch approaches to the motion and structure problem (for recent results see [10, 22, 24]).

Gennery [7, 8] considers the problem of tracking 3-D objects of known shape, and recovering their motion and position with respect to the observer, his state vector therefore consists of 12 parameters (6 configuration and 6 velocity); he treats both point and line features. Matthies et al. [17] use a filter formulation similar the the one described in section 2. They make the assumption of known motion, and demonstrate their scheme in the special case of translational motion both with a feature-based scene representation and

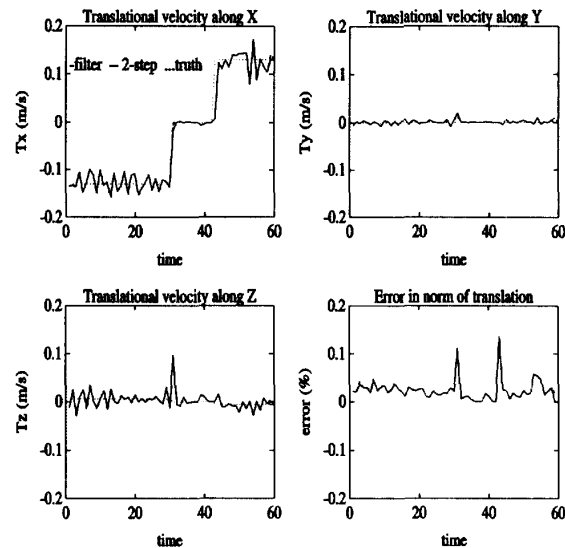


Figure 2: The motion of the observer with respect to his own reference is rotational about the vertical axis with velocity of $\theta = 5^\circ/\text{step}$, while translational velocity is $1 - \cos(\theta)$ along Z and $\sin(\theta)$ along Y. After 30 steps the observer stop translating and keeps on rotating. After 6 more steps he inverts the direction of rotation and eventually, after 42 steps, he starts translating back to his original position. Zero-mean Gaussian noise was added to the observations with a standard deviation corresponding to 1 pixel. In this figure we show plots of the estimation of translation: in the top-left plot the X component of the translational velocity is shown: the estimate of the filter (solid) is compared with the ground truth (dotted) and the 2-views algorithm (dashed). As it can be seen the X component of motion presents two first order discontinuities, which challenge the dynamical model used for motion and for the scale factor. Notice that the estimate keeps within 2% standard deviation error (bottom-right plot). In order to keep the motion model flexible its variance is kept relatively large (10^{-2}), hence the filter tracks closely the 2-views algorithm. Smaller variance values would result in more smoothing, but also poor handling of motion discontinuities.

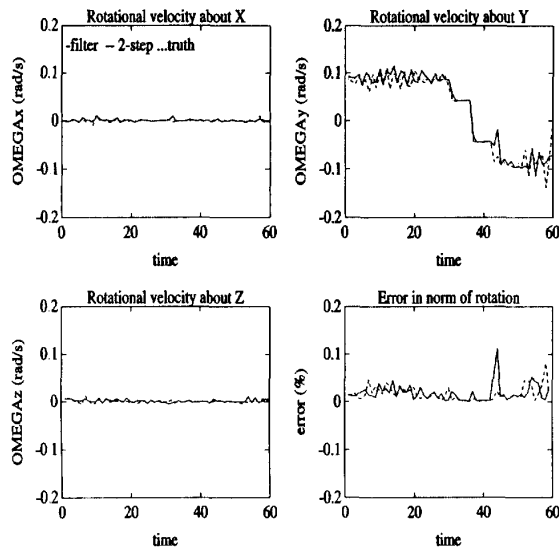


Figure 3: Estimation of rotational velocity. Notice that between time steps 30 and 42 translation is zero, while rotation is non-zero and presents first order discontinuities. However the filter is still able to take advantage of the reliable measurements of rotation and estimate it correctly. Also it keeps updating the structure which is changing due to non-zero rotation, exploiting the dynamical state equation (rigidity assumption). The bottom-right plot shows the rotation estimation error, which is within 2% standard deviation.

with a dense depth-map. Recently Shekhar and Chellappa have successfully tested a similar approach, again assuming known motion [20]. Heel [11, 12, 13] proposed a sequel of filter formulations for the recovery of both structure and motion, with further regularization of the depth map. The essence of the formulation is similar to the one by Matthies et al., but the motion is now estimated at each step with a least squares procedure from the estimated structure: the best estimate of motion is used to produce an estimate for structure which is used to obtain the new estimate of motion. The filter has adaptive structure and shows severe convergence and initialization problems (convergence has been experimentally demonstrated for 3 degrees of freedom, see [12] pag. 38). In Heel's scheme the motion parameters are not included in the state of the filter, hence motion temporal coherence cannot be exploited.

None of the above approaches describes and discusses error dynamics. We suspect that most schemes keep the error variances constant in time. As a consequence the filter needs to be tuned for each individual motion condition and only constant motions are handled correctly. This might be the reason why in the literature one finds that simulations and testing are only performed for very simple motions. To our knowledge no current approach performs a complete and

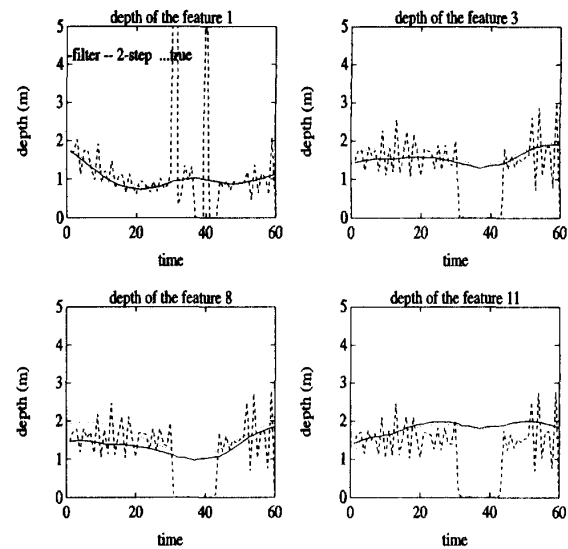


Figure 4: Depth for some points chosen at random among the 20 observed: The performance of the filter (solid) is compared with that of the two-frames step (dashed). Ground truth is the dotted line. As it can be seen, during the period in which translation is zero, the two-frames step is unable to resolve depth and feeds the filter with meaningless measurements. However it also warns the filter that those data have very high variance, so that the gain saturates selectively along the components of translation and depth. On the contrary measurements of rotation and image coordinates are still reliable, and the filter exploits the model equation and the still reliable measurements of rotation to update structure. Notice the transient of approximately 20 time-steps.

rigorous treatment of measurement and model errors.

5 Conclusions

In this paper we have presented a recursive algorithm to recover both the structure of a scene and ego-motion from a monocular sequence of images of a set of point features. The algorithm is based on an Extended Kalman Filter cascaded to a 2-perspective-views motion and structure step. The 2-views step returns to the filter instantaneous estimates of motion and structure and variances of such estimates. The filter has as its input all of the components of the state, hence the model is globally observable and optimally conditioned with respect to motion.

We have demonstrated the features of our algorithm in a simulation with a motion pattern including (a) discontinuities and (b) ill-conditioned tracts. The simulations show that (a) discontinuities are correctly handled, i.e. the filter is not based on excessive smoothness assumptions, (b) depth is estimated in a robust fashion with respect to the motion:

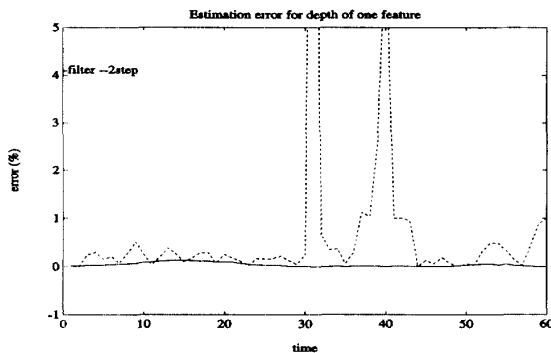


Figure 5: Error in the determination of depth: the performance of the filter (solid) is compared with that of the two-frames step (dashed). As it can be seen as the observer stops translating the 2-frames algorithm can no longer resolve depth, however the filter discards the unreliable data and keeps updating structure based on the state dynamical equation and the still reliable measurements of rotation.

when translation drops to zero the measurements from the 2-frames algorithm are inaccurate, hence the filter discards them and estimates depth correctly using the state model.

Our algorithm has a simple formulation, it is robust with respect to motion singularities and general since it requires only visual input (and one initial scale parameter if one is to recover the scale factor). Furthermore the formulation is complete since the only unknown parameters are the variances of the model error, for which the tuning procedure is carried out. In order to recover the scale factor we need some external information. Once this information is available at one step it is propagated to subsequent frames allowing recovery of the scale factor throughout the sequence. Additional information about the motion dynamics, if available, may be inserted into the state of the filter.

The filter is robust with respect to variation of the tuning parameters.

The computational cost of one step of the filter is approximately $O(N^3)$, where N is the number of observed points, since the filter performs matrix additions, multiplications and one inversion of the variance matrix for the pseudo-innovation.

We have not yet addressed a number of issues: (a) The algorithm has not been tested on real image sequences. (b) Use of filter-generated predictions on feature positions and variances in the first module (feature tracking/optical flow) could be added for improving efficiency (shown as dashed-line outer loop in fig. 1). This method is used in road-following vehicle navigation experiments by Dickmanns et al. [5]. (c) Instead of the features position measured in the $t - 1$ image one could use the best current estimate of the position of the features from the filter in the 2-views algorithm (shown as dashed-line middle loop in fig. 1). (d) Integration with stereo is a natural extension. (e) Numerical issues should

also be explored to improve computational efficiency; for example how to exploit the structure of the matrix Q through the Riccati equation in computing the gain of the filter.

References

- [1] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press, 1987.
- [2] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. RPL-TR 9107, Queen's University Kingston, Ontario, Robotics and perception laboratory, 1992. Also in Proc. CVPR 1992, pp 236-242.
- [3] T. Broida and R. Chellappa. Estimating the kinematics and structure of a rigid object from a sequence of monocular frames. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1991.
- [4] T. Broida and R. Chellappa. Estimation of object motion parameters from noisy images. *IEEE Trans. Pattern Anal. Mach. Intell.*, Jan. 1986.
- [5] E. D. Dickmanns and V. Graefe. Applications of dynamic monocular machine vision. *Machine Vision and Applications*, 1:241-261, 1988.
- [6] O. Faugeras and S. Maybank. Motion from point matches: multiplicity of solutions. *Int. J. of Computer Vision*, 4:225-246, 1990.
- [7] D.B. Gennery. Tracking known 3-dimensional object. In *Proc. AAAI 2nd Natl. Conf. Artif. Intell.*, pages 13-17, Pittsburg, PA, 1982.
- [8] D.B. Gennery. Visual tracking of known 3-dimensional object. *Int. J. of Computer Vision*, 7(3):243-270, 1992.
- [9] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. 2nd Europ. Conf. Comput. Vision*, G. Sandini (Ed.), LNCS-Series Vol. 588, Springer-Verlag, 1992.
- [10] D.J. Heeger and A.D. Jepson. Subspace methods for recovering rigid motion i. *Int. J. of Computer Vision*, 7(2):95-117, 1992.
- [11] J. Heel. Dynamic motion vision. *Robotics and Autonomous Systems*, 6(1), 1990.
- [12] J. Heel. Direct estimation of structure and motion from multiple frames. *AI Memo 1190, MIT AI Lab*, March 1990.
- [13] J. Heel. Temporal integration of 3-d surface reconstruction. To appear on *IEEE trans. PAMI, special issue on the interpretation of 3-D scenes.*, March 1991.
- [14] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
- [15] R.E. Kalman. A new approach to linear filtering and prediction problems. *Trans. of the ASME-Journal of basic engineering.*, 35-45, 1960.
- [16] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133-135, 1981.
- [17] L. Matthies, R. Szelisky, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *Int. J. of computer vision*, 1989.
- [18] P. Perona and S. Soatto. Motion and structure from 2 perspective views of p localized features: noise sensitivity analysis. Technical report, California Institute of Technology, 1992.
- [19] J. Philip. Estimation of three dimensional motion of rigid objects from noisy observations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(1):61-66, 1991.
- [20] C. Shekhar and R. Chellappa. Passive ranging using a moving camera. *J. of Robotics S. vol. 9*, 1992.
- [21] S. Soatto and P. Perona. Temporal integration of motion information for the robust recovery of motion and structure. Technical report, California Institute of Technology, 1992.
- [22] M.E. Spetsakis and Y. Aloimonos. A multiframe approach to visual motion perception. *Int. J. of Computer Vision*, 6:245-255, 1991.
- [23] M.E. Spetsakis and Y. Aloimonos. Optimal visual motion estimation: a note. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14:959-964, 1992.
- [24] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. J. of Computer Vision*, 9(2):137-154, 1992.
- [25] R.Y. Tsai and T.S. Huang. *Image understanding*, S. Ullman and W. Richards (eds.), chapter Uniqueness and estimation of three-dimensional motion parameters of rigid objects. Ablex Publishing, Norwood NJ, 1984.
- [26] 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.