April, 1993 LIDS-P 2174

Research Supported By:

ONR grant N00014-91-J-1004
ONR grant N00014-91-J-1120
ARO grant DAAL03-92-G-0115
NSF grant MIP-9015281

Probabilistic and Sequential Computation of Optical Flow Using Temporal Coherence* (revised version of LIDS-P-2122)

Chin, T.M.
Karl, W.C.
Willsky, A.S.

LIDS-P-2174
(revised version of
LIDS-P-2122)

Probabilistic and Sequential Computation of Optical Flow using Temporal Coherence[1]

Toshio M. Chin[2] William C. Karl[3] Alan S. Willsky[4]

revised April 12, 1993

[1]This research was supported in part by the Office of Naval Research under Grants N00014-91-J-1004 and N00014-91-J-1120 (Random Field in Oceanography ARI), the National Science Foundation under Grant MIP-9015281, and the Army Research Office under Grant DAAL03-86-K-0171. Address for correspondence: Toshio Chin, RSMAS-MPO, 4600 Rickenbacker Causeway, Miami, FL 33149.

[2]with the Rosenstiel School of Marine and Atmospheric Science, University of Miami, and formally with the Laboratory for Information and Decision Systems, M.I.T.

[3]with the Laboratory for Information and Decision Systems, M.I.T.

[4]with the Laboratory for Information and Decision Systems and the Department of Electrical Engineering and Computer Science, M.I.T.

Abstract

In the computation of dense optical flow fields, spatial coherence constraints are commonly used to regularize otherwise ill-posed problem formulations, providing spatial integration of data. In this paper we present a temporal, multi-frame extension of the dense optical flow estimation formulation proposed by Horn and Schunck [1] in which we use a *temporal coherence constraint* to yield the optimal fusing of data from multiple frames of measurements. Conceptually at least, standard Kalman filtering algorithms are applicable to the resulting multi-frame optical flow estimation problem, providing a solution which is sequential and recursive in time. Experiments are presented which demonstrate that the resulting multi-frame estimates are more robust to noise than those provided by the original, single-frame formulation. Also, we demonstrate cases where the aperture problem of motion vision cannot be resolved satisfactorily without the temporal integration of data enabled by the proposed formulation. Practically, the large matrix dimensions involved in the problem prohibit exact implementation of the optimal Kalman filter. To overcome this limitation we present a computationally efficient, yet near optimal approximation of the exact filtering algorithm. This approximation has a precise interpretation as the sequential estimation of a reduced order spatial model for the optical flow estimation error process at each time step and arises from an estimation-theoretic treatment of the filtering problem. Experiments also demonstrate the efficacy of this near optimal filter.

Contents

1 Introduction

Computation of the dense, 2-D vector field of apparent motion, or *optical flow (image flow)*, is of considerable interest in image sequence processing. It is an important "low-level" step in many of the hierachical approaches to computational vision — both for the development of artificial visual systems in robotics and for the modeling of biological visual systems. For example, optical flows can provide us with the information necessary to detect object boundaries [2, 3] and to derive the 3-D motion and structure of the objects in an image frame [4, 5, 6]. Optical flow computation is also important to applications in fields outside of robotics and cognitive sciences, such as in assessing motility of the heart [7, 8] and in interpretation and prediction of marine and atmospheric processes [9, 10, 11]. Motion information is additionally useful for managing the image sequences themselves, as it offers a basis for image sequence compression for efficient transmission and storage [12, 13]. There exist a variety of techniques for computing optical flow, including [1, 14, 15, 16, 17, 18, 19, 20, 21], as well as comparative studies of them [22, 23]. The focus of this paper is *not* to provide a fundamentally new method of optical flow computation but to study, in a probabilistic framework, how the flow estimates can be improved by incorporating a longer sequence of images and how to compute such improved estimates in a computationally efficient and near-optimal manner.

In the computation of a dense optical flow field at a single point in time, spatial coherence (smoothness) constraints are commonly used to regularize an otherwise ill-posed formulation by the spatial integration of data. Spatial coherence, in a sense, represents our prior knowledge or assumption that the moving object (solid, viscous fluid, etc.) in the scene is structurally integral and smooth. In this paper we examine the effects of applying a similar constraint over time. In particular, we present a temporal extension of the formulation proposed by Horn and Schunck [1] in which we use a *temporal coherence constraint* captured by an evolution equation to provide the optimal fusing of data from multiple frames of measurements. Estimating the optical flow field by processing *sequences* of measurements has an obvious advantage over static estimation based on only a single such observation. For one thing, the accumulation of a larger quantity of data leads to a more reliable estimate due to a reduction in measurement noise. Another advantage, not as obvious, is that in some cases a single measurement may not provide sufficient information

to resolve static ambiguities in the flow field (i.e. the aperture problem of computational vision [24]) and hence for reasonable estimates to be obtained, temporal information must be utilized as well. Such ambiguity is caused by a lack of spatial diversity in the direction of the spatial gradient [1]. In many cases the desired diversity of gradient directions *is* available over *time*, allowing the resolution of this ambiguity through the incorporation of more image frames, as exemplified in this paper.

The spatial and temporal coherence constraints can be interpreted as *a priori* statistical description of the unknown field [25]. Specifically, the optical flow formulation by Horn and Schunck can be considered as a Bayesian estimation problem with additive Gaussian noise. Utilizing such an estimation-theoretic framework for optical flow computation, we model the time-varying unknown flow field as a dynamic process with an associated evolution equation which captures the temporal coherence constraint of the variational formulation. Thus, the temporal extension of the Horn and Schunck formulation, i.e., multi-frame optical estimation, can be written as state estimation based on a dynamic system to which well-developed optimal sequential estimation algorithms, such as the Kalman filter and its variants, are applicable for solution [26, 25]. The probabilistic framework allows quantification of uncertainty in the estimates through computation of the second order statistics.

Although Kalman filter allows time-recursive estimation of multi-frame optical flow fields, its computational demands are still impractical. For typical problems the dimension of the associated state will be on the order of the number N of pixels in the image, typically 10^4 to 10^6 elements. The associated covariance matrices, necessary for implementation of the optimal filter, will thus have on the order of 10^8 to 10^{12} elements! The storage and manipulation of such large matrices is clearly prohibitive, necessitating the use of a suboptimal method. Our model-based approach provides a rational basis for the design of a computationally feasible, yet nearly optimal filter for optical flow estimation which naturally incorporates both temporal and spatial coherence constraints. Our approximate filter arises from the construction of a *reduced order spatial model* for the optical flow estimation error field at each time step and thus possesses a precise estimation-theoretic interpretation. Reduced order approximations for Kalman filtering on (2-D) image frames have frequently been reported [27, 28, 29]; the computational algorithm in this paper represents a 3-D version for image *sequences* in which we must determine a reduced-order model, at each frame in the

image sequence, in order to capture the dynamically-evolving statistical structure of the estimation error field.

The paper is organized as follows. In Section 2 we review the classical single-frame optical flow estimation problem in a continuous setting. In Section 3 we present our temporal extension to the continuous classical problem. In Section 4, we give a discrete reformulation of the single-frame formulation and its interpretation as a maximum likelihood estimation problem. We then derive a statistically optimal Kalman filtering algorithm for the resulting multi-frame problem. In Section 5, we investigate implementation issues, including approximation of the Kalman filter and the effects of discretization on the fundamental measurement constraint. In Section 6, we present various simulation results demonstrating the benefits of applying temporal coherence to multi-frame optical flow estimation as well as the effectiveness of our approximate Kalman filter in computing such flow fields. The paper concludes with final comments in Section 7.

2 Optical Flow Estimation

We perceive motion by temporally tracking image intensity patterns, often associated with reflections from the surfaces of objects in the scene. If the brightness corresponding to a point on the object surface remains practically constant for a sufficiently long duration, the position of the point can be tracked by referencing the same brightness value, leading to motion perception. Such an assumption of brightness invariance can be expressed as [1]

$$\frac{d}{dt}E = 0 \tag{1}$$

where $E(s_1, s_2, t)$ is the image intensity, which is treated as a differentiable scalar function over the image frame $(s_1, s_2) \subset \mathcal{D}$ and time t. Let $f(s_1, s_2, t) \equiv \left[\frac{ds_1}{dt}, \frac{ds_2}{dt}\right]^T$ be the optical flow vector at a given point in the image frame and time. By expanding (1) in terms of partial derivatives, we obtain the following relationship between the image intensity gradients and the optical flow vector at each point in the space-time domain:

$$\frac{\partial E}{\partial t} + \left[\frac{\partial E}{\partial s_1}, \frac{\partial E}{\partial s_2}\right] f = 0. \tag{2}$$

Horn and Schunck [1] have suggested that the fact that (2) provides only one constraint for the two unknown components of f is the reason for the visual ambiguity often referred to as the *aperture problem* in psychophysics [24] and have provided a method to compute optical flow using additional constraints. Their method of computing the optical flow finds a single frame of the flow field, i.e. $f(s_1, s_2, t)$ for a fixed t, as a the solution of a quadratic minimization problem

$$\min_{f(s_1,s_2,t)} \iint_{\mathcal{D}} \nu \left\| \frac{\partial E}{\partial t} + \left[\frac{\partial E}{\partial s_1}, \frac{\partial E}{\partial s_2} \right] f \right\|^2 + \mu_1 \left\| \frac{\partial}{\partial s_1} f \right\|^2 + \mu_2 \left\| \frac{\partial}{\partial s_2} f \right\|^2 ds_1\, ds_2 \tag{3}$$

where $\nu(s_1, s_2, t) \neq 0$, μ_1, and μ_2 are given weights. The first quadratic term involves the image data, penalizing large deviations from (2). The second and third terms are necessary to make the formulation mathematically well-posed [30]. These two terms also represent our prior belief about the flow field, implying that the computed flow should vary smoothly over space. Such *spatial coherence* of the flow vectors reflects the smoothness and stiffness of the object surface in the scene [24].

3 Multi-frame Formulation

The formulation (3) processes the data (the gradients of the image intensity) a frame at a time, yielding flow estimates independently over time. We now consider the imposition of *temporal coherence* [31] to the flow field in addition to the spatial coherence enforced by (3) in order to utilize more data for each flow vector estimate. Temporal coherence imposes an inertia condition on the flow field, favoring gradual changes in the optical flow vectors over time. Temporal coherence models of optical flow are applicable to a wide range of motions in natural scenes, as most movements display inertia of some type. To obtain such a *multi-frame* formulation of the optical flow computation problem we use a simple temporal extension [25, 26] of (3). In particular, for $0 \leq t \leq \tau$ we find the flow field $\widehat{f}(s_1, s_2, t)$ which provides the solution to

$$\min_{f(s_1,s_2,t)} \int_0^\tau \iint_{\mathcal{D}} \nu \left\| \frac{\partial E}{\partial t} + \left[\frac{\partial E}{\partial s_1}, \frac{\partial E}{\partial s_2} \right] f \right\|^2 + \mu_1 \left\| \frac{\partial}{\partial s_1} f \right\|^2 + \mu_2 \left\| \frac{\partial}{\partial s_2} f \right\|^2 + \rho \left\| \frac{\partial}{\partial t} f \right\|^2 ds_1\, ds_2\, dt. \tag{4}$$

The multi-frame formulation (4) is obtained from the single-frame formulation (3) by the addition of a quadratic term involving the first order temporal derivative.

We demonstrate in Section 6 that flow estimates based on this multi-frame problem extension, arising from the imposition of temporal coherence, yields solutions which are more robust and accurate than those derived using the corresponding single-frame formulation (3). The inclusion of the temporal constraint in (4) allows the integration of data over time. This use of temporally extended data can both help resolve single-frame observation ambiguities (the aperture problem) and greatly reduce noise sensitivity relative to the non-temporal formulation (3).

Note that the full solution to the optimization problem (4) leads to a space-time field $\widehat{f}(s_1, s_2, t)$ in which the solution at any time takes advantage of all available constraints over the entire time interval $0 \leq t \leq \tau$. In the parlance of estimation theory, this is the noncausal, *smoothed* estimate. In this paper we focus on the *filtered* estimate, i.e. the value of the solution to (4) at the current time $t = \tau$. Filtered estimates are desirable in applications where the optical flow needs to be calculated as soon as each frame in the image sequence becomes available; however, obtaining such estimates corresponds to solving a *different* 3-D optimization problem for each τ as τ increases. Such a solution clearly results in a greatly increased computational burden over what is required for the single-frame solution of (3), making direct solution of the optimization problem (4) prohibitive.

We present an efficient, recursive solution to a discrete and probabilistic version of the causal filtering problem represented by (4), obtained through Kalman filtering. Applications of Kalman filtering to various formulations of optical flow estimation [32, 33] as well as to other low-level reconstruction problems in computational vision [25] have been proposed. In these works, simplified formulations are used to avoid the computational complexity associated with filtering large amounts of image data. Specifically, the uncertainty in the dynamic model for the time-varying unknown field, and hence the uncertainty in the estimate itself, is not formally represented or properly propagated in these approaches. In an exact implementation of a Kalman filter, such uncertainty, as captured in the estimation error covariance matrix, is propagated along with the estimate itself [34, 35, 36] and allows for the optimal fusing of the current estimate with new observations. The filtering algorithm presented in this paper employs a more systematic and rational approximation of Kalman filter than those previously reported; it is based on the propagation of approximate local *models* of the estimation error covariance. The mathematical details of our approximation techniques can be found in [26] in the more general context of low-level visual reconstruction.

4 Discretization and Probabilistic Interpretation

4.1 Single-Frame Case

To obtain a discrete formulation of the single-frame problem (3) we sample the image frame $\mathcal{D}$ on an $n_1 \times n_2$ rectangular grid containing $N \equiv n_1 n_2$ points. Let $\mathbf{f}(t)$ be a vector of the flow values $f(s_1, s_2, t)$ sampled on the grid and ordered lexicographically according to the sampled spatial coordinates (s_1, s_2). Since f is a 2-vector at each point, $\mathbf{f}(t)$ has $2N$ elements. We similarly define $\mathbf{g}(t)$ to be an N-vector of the lexicographically ordered samples of $(-\frac{\partial E}{\partial t})$ taken on the same grid. Finally, let $\mathbf{H}(t)$ and $\mathbf{W}(t)$ be block diagonal matrices whose diagonal elements are the samples of $\left[\frac{\partial E}{\partial s_1}, \frac{\partial E}{\partial s_2}\right]$ and $\nu(s_1, s_2, t)$, respectively, taken on the same grid and sequenced in the same lexicographical order. A discrete version of the single-frame formulation (3) is then given by

$$\min_{\mathbf{f}(t)} \left\{ \|\mathbf{g}(t) - \mathbf{H}(t)\mathbf{f}(t)\|^2_{\mathbf{W}(t)} + \|\mathbf{S}_1\mathbf{f}(t)\|^2_{\mu_1 I_{2(N-n_2)}} + \|\mathbf{S}_2\mathbf{f}(t)\|^2_{\mu_2 I_{2(N-n_1)}} \right\} \tag{5}$$

where $\|\mathbf{x}\|^2_{\mathbf{M}}$ denotes the weighted norm $\mathbf{x}^T\mathbf{M}\mathbf{x}$, I_m represents the $m \times m$ identity matrix, and $\mathbf{S}_1$ and $\mathbf{S}_2$ are first order difference operators along the s_1 and s_2 axes, respectively, given by

$$\mathbf{S}_1 \equiv \begin{bmatrix} D_{2(n_1-1)} & & \\ & \ddots & \\ & & D_{2(n_1-1)} \end{bmatrix}, \quad D_{2(n_1-1)} \equiv \begin{bmatrix} -I_2 & I_2 & & \\ & \ddots & \ddots & \\ & & -I_2 & I_2 \end{bmatrix},$$

$$\mathbf{S}_2 \equiv \begin{bmatrix} -I_{2n_1} & I_{2n_1} & & \\ & \ddots & \ddots & \\ & & -I_{2n_1} & I_{2n_1} \end{bmatrix}.$$

Solving the quadratic minimization problem (5) is equivalent to solving a maximum likelihood estimation problem [36] for $\mathbf{f}(t)$ with the following *observation equation*:

$$\begin{bmatrix} \mathbf{g}(t) \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{H}(t) \\ \mathbf{S}_1 \\ \mathbf{S}_2 \end{bmatrix} \mathbf{f}(t) + \mathbf{r}(t), \quad \mathbf{r}(t) \sim \left(0, \begin{bmatrix} \mathbf{W}^{-1}(t) & & \\ & \mu_1^{-1} I_{2(N-n_2)} & \\ & & \mu_2^{-1} I_{2(N-n_1)} \end{bmatrix} \right) \tag{6}$$

where we have used the notation $\mathbf{x} \sim (\mathbf{m}, \mathbf{C})$ to denote a Gaussian random vector $\mathbf{x}$ whose mean and covariance are $\mathbf{m}$ and $\mathbf{C}$, respectively. Thus, $\mathbf{r}(t)$ is a zero-mean Gaussian random noise process. Note that $\mathbf{W}^{-1}(t)$ is a diagonal covariance matrix, whose non-zero elements $\nu^{-1}(s_1, s_2, t)$ are variances representing probabilistically how much the measured image gradients deviate from the ideal brightness constraint (2). Also, μ_1^{-1} and μ_2^{-1} are the variances representing how much the first-order differences between neighboring flow vectors deviate from zero, effectively controlling the strength of the spatial coherence constraint. The maximum likelihood estimate for the optical flow, $\widehat{\mathbf{f}}(t)$, is obtained as the solution of the following inverse problem:

$$\left(\mathbf{H}^T(t)\mathbf{W}(t)\mathbf{H}(t) + \mu_1\mathbf{S}_1^T\mathbf{S}_1 + \mu_2\mathbf{S}_2^T\mathbf{S}_2\right)\widehat{\mathbf{f}}(t) = \mathbf{H}^T(t)\mathbf{W}(t)\mathbf{g}(t). \tag{7}$$

The equations in (7) represent a discrete version of the coupled Poisson equations of the Horn and Schunck formulation. Note that the matrix operator $\mathbf{L}(t) = (\mathbf{H}^T(t)\mathbf{W}(t)\mathbf{H}(t) + \mu_1\mathbf{S}_1^T\mathbf{S}_1 + \mu_2\mathbf{S}_2^T\mathbf{S}_2)$ on the left hand side of (7) has a sparse, *nearest neighbor* (a nested block tri-diagonal) structure [37], whose sparseness enables us to use efficient iterative procedures, such as multigrid methods [38], in the solution of (7). Also, this sparse matrix corresponds to the *information matrix* (the inverse of the covariance matrix) associated with the posterior estimation error $\mathbf{d}(t) \equiv \mathbf{f}(t) - \widehat{\mathbf{f}}(t)$. In particular $\mathbf{L}(t)$ can naturally be considered to specify an implicit *model* for the estimation error process $\mathbf{d}(t)$ of the following form:

$$\mathbf{L}(t)\,\mathbf{d}(t) = \zeta(t), \qquad \zeta(t) \sim (\,0, \mathbf{L}(t)\,). \tag{8}$$

The nearest neighbor structure of $\mathbf{L}(t)$ in (8) or (7) reflects a corresponding local structure to the statistical model for the estimation error covariance. We will use this observation in what follows.

4.2 Multi-Frame Case

Now we consider the multi-frame extension of the single-frame formulation given in (6). The continuous optimization problem (4) can be considered to be an *optimal smoothing* problem based on the following temporal, linear Gauss-Markov dynamic system for $\mathbf{f}(t)$ [26]:

$$\frac{\partial}{\partial t} f(s_1, s_2, t) = q(t) \tag{9}$$

where $q(t)$ is a Gaussian white noise process of zero mean and intensity ρ^{-1}. For such an optimal smoothing problem, *two-filter* methods (obtained by running a Kalman filter in each of the causal and anti-causal directions [36]) are applicable [39, 40]. In general we wish to compute only the most recent estimate $\widehat{\mathbf{f}}(s_1, s_2, \tau)$ from (4) for each $\tau \geq 0$. Such an estimate can be obtained by a single causal Kalman filter. Specifically, a *discrete* version of this multi-frame problem can be formulated as a state estimation problem for the dynamic system whose dynamic equation is

$$\mathbf{f}(t) = \mathbf{f}(t-1) + \mathbf{q}(t), \quad \mathbf{q}(t) \sim \left(0, \rho^{-1} I\right), \tag{10}$$

and whose observation equation is given by (6). The process noise $\mathbf{q}(t)$ is uncorrelated over time and captures the uncertainty in the dynamic model (10). This Gauss-Markov dynamic model, a discrete version of (9), indicates that the optical flow evolves in time as the accumulation of a random perturbation at each time frame. In this paper we will be primiarily concerned with the temporal dynamics of the form (10). Naturally, other more complicated dynamic models, corresponding to having different temporal coherence terms in (4), could be used in general.

State estimation for the dynamic system (10)(6) can be performed in a sequential way by the following implementation of the *information form* [34, 36] of the Kalman filter:

- prediction stage

$$\overline{\mathbf{L}}(t) = \rho I - \rho^2 \left(\widehat{\mathbf{L}}(t-1) + \rho I\right)^{-1} \tag{11}$$

$$\overline{\mathbf{f}}(t) = \widehat{\mathbf{f}}(t-1) \tag{12}$$

$$\overline{\mathbf{z}}(t) = \overline{\mathbf{L}}(t)\overline{\mathbf{f}}(t) \tag{13}$$

- update stage

$$\widehat{\mathbf{L}}(t) = \overline{\mathbf{L}}(t) + \mathbf{H}^T(t)\mathbf{W}(t)\mathbf{H}(t) + \mu_1 \mathbf{S}_1^T \mathbf{S}_1 + \mu_2 \mathbf{S}_2^T \mathbf{S}_2 \tag{14}$$

$$\widehat{\mathbf{z}}(t) = \overline{\mathbf{z}}(t) + \mathbf{H}^T(t)\mathbf{W}(t)\mathbf{g}(t) \tag{15}$$

$$\widehat{\mathbf{L}}(t)\widehat{\mathbf{f}}(t) = \widehat{\mathbf{z}}(t) \tag{16}$$

where $\overline{\mathbf{f}}(t)$ is the one-step predicted estimate and $\widehat{\mathbf{f}}(t)$ is the updated estimate using the new data

available at time t. Also, $\overline{\mathbf{L}}(t)$ and $\widehat{\mathbf{L}}(t)$ denote the predicted and updated information matrices, respectively. Note that the updated estimate $\widehat{\mathbf{f}}(t)$ in (16) is specified implicitly, as for the single-frame case (7).

5 Implementational Issues

5.1 Suboptimal Kalman filtering

The number of pixels, N, in a frame of a typical image sequence is on the order of 10^4 to 10^6. Such a large number of points makes direct implementation of the optimal information Kalman filter (11)–(16) impractical as the associated information matrices $\overline{\mathbf{L}}(t)$ and $\widehat{\mathbf{L}}(t)$ of the optimal filter will have on the order of 10^8 to 10^{12} elements.

To understand our sub-optimal filter, consider the set of equations (11)-(16). First consider the update stage of the Kalman filter. If $\overline{\mathbf{L}}(t)$ possesses a sparse and banded nearest neighbor structure, as was true for the single-frame problem, then (14) will preserve this structure in $\widehat{\mathbf{L}}(t)$ since, as we pointed out in connection with (7), $\mathbf{H}^T(t)\mathbf{W}(t)\mathbf{H}(t) + \mu_1\mathbf{S}_1^T\mathbf{S}_1 + \mu_2\mathbf{S}_2^T\mathbf{S}_2$ also possesses this structure. In particular, if this is the case, then (16) may still be solved efficiently for the updated estimate $\widehat{\mathbf{f}}(t)$, and in fact this step would have *exactly* the same computational complexity as in the single-frame approach. Thus we desire to preserve such a sparse and banded structure in $\overline{\mathbf{L}}(t)$.

Now consider the prediction stage. Unfortunately, even if $\widehat{\mathbf{L}}(t-1)$ in (11) is initially sparse and banded, the predicted information matrix $\overline{\mathbf{L}}(t)$ will not be due to the matrix inverse on the right hand side of this equation. In addition, finding the inverse of this matrix is a prohibitively complex procedure. What we desire, then, is a sparse and banded ***approximation*** to $\overline{\mathbf{L}}(t)$ that may be efficiently computed.

As detailed in [26, 41], such an approximation may indeed be obtained by expanding the matrix inverse on the right hand of (11) in a series as follows:

$$\overline{\mathbf{L}}(t) \;=\; \rho I - \rho^2(\Lambda^{-1} - \Lambda^{-1}\Omega\Lambda^{-1} + \Lambda^{-1}\Omega\Lambda^{-1}\Omega\Lambda^{-1} - \cdots) \tag{17}$$

where Λ is a block diagonal matrix whose 2×2 diagonal blocks are identical to the corresponding diagonal blocks of the matrix $\widehat{\mathbf{L}}(t-1)+\rho I$ while $\Omega \equiv \widehat{\mathbf{L}}(t-1)+\rho I - \Lambda$ is given by the remaining off-

diagonal part of $\hat{\mathbf{L}}(t-1)+\rho I$. Note that Λ^{-1} is block diagonal. The series (17) may now be truncated to any desired number of terms to obtain an approximation to the exact expression of the desired level of accuracy. The more terms are kept, the less sparse and banded the approximation will become. Thus, there is a tradeoff between accuracy and computational efficiency. Our experience has shown that retaining only the first two terms yields excellent results. In particular, we obtain our near-optimal filter by replacing the optimal prediction step (11) by the following two-term approximation:

$$\overline{\mathbf{L}}(t) \;=\; \rho I - \rho^2(\Lambda^{-1} - \Lambda^{-1}\Omega\Lambda^{-1}). \tag{18}$$

Unlike (11), the suboptimal prediction step (18) does indeed preserve the desired nearest neighbor structure in the (approximated) information matrix $\overline{\mathbf{L}}(t)$.

It can be verified straightforwardly that propagating the information matrix in the approximate filter as in (14) and (18) costs only $O(N)$ flops per frame and has a local, modular computational structure suitable for parallel implementation. Throughout the filtering procedure, the approximated information matrices maintain the nearest neighbor structure and have only $O(N)$ non-zero elements. Thus, the approximate filter has significant computational and storage advantages over the optimal Kalman filter, which normally requires $O(N^2)$ storage elements and $O(N^3)$ flops per frame of data.

A useful way to understand our approximation is provided by an examination of the update stage of the Kalman filter. In this part of the filter we are fusing the information from the previous prediction stage, as captured by $\overline{\mathbf{L}}(t)$ and $\overline{\mathbf{z}}(t)$ (or equivalently $\overline{\mathbf{f}}(t)$), with the new observation. In particular, $\overline{\mathbf{L}}(t)$ can naturally be thought of as specifying a prior model for the error $\mathbf{e}(t) \equiv \mathbf{f}(t)-\overline{\mathbf{f}}(t)$ in the current estimate of the following form:

$$\overline{\mathbf{L}}(t)\,\mathbf{e}(t) = \zeta(t), \qquad \zeta(t) \sim \Big(\,0, \overline{\mathbf{L}}(t)\,\Big) \tag{19}$$

which is just the counterpart of (8) for the dynamic problem. This model is then combined with the new observation to produce the best estimate $\hat{\mathbf{e}}(t)$ of this error. The updated estimate $\hat{\mathbf{f}}(t)$ in (16) is then equal to $\overline{\mathbf{f}}(t) + \hat{\mathbf{e}}(t)$. The update stage is thus just a static spatial estimation problem, where (19) represents a prior model just before the inclusion of new data. That is, by writing the

observation equation (6) as $\mathbf{y}(t) = \mathbf{C}(t)\mathbf{f}(t) + \mathbf{r}(t)$ (where $\mathbf{y}(t)$ and $\mathbf{C}(t)$ have been introduced to denote the corresponding vector and matrix in (6) concisely), the estimate $\widehat{\mathbf{e}}(t)$ can be obtained by solving the static estimation problem

$$\begin{bmatrix} 0 \\ \mathbf{y}(t) - \mathbf{C}(t)\bar{\mathbf{f}}(t) \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{L}}(t) \\ \mathbf{C}(t) \end{bmatrix} \mathbf{e}(t) + \begin{bmatrix} -\zeta(t) \\ \mathbf{r}(t) \end{bmatrix}, \tag{20}$$

which is statistically equivalent to obtaining the updated estimate $\widehat{\mathbf{f}}(t)$ of the unknown $\mathbf{f}(t)$ given the prediction $\bar{\mathbf{f}}(t)$. Since the implicit model is specified by $\bar{\mathbf{L}}(t)$, our approximation of this matrix by a sparse matrix of the given nearest neighbor structure in (18) corresponds naturally to the specification of an approximate, *reduced-order* model for the error process. In particular this approximation may be viewed as the imposition of a Markov Random Field structure of fixed spatial extent on the flow field estimation-error [26]. Our approximation thus has a rational basis in estimation-theoretic considerations.

5.2 Variance computation

The estimation error covariance matrix $\widehat{\mathbf{P}}(t)$ associated with the updated flow estimate $\widehat{\mathbf{f}}(t)$ is the inverse of the information matrix $\widehat{\mathbf{L}}(t)$. This inversion can be performed recursively as

$$\widehat{\mathbf{P}}^{(k+1)}(t) = \Lambda_L^{-1} - \Lambda_L^{-1}\Omega_L\widehat{\mathbf{P}}^{(k)}(t) \tag{21}$$

where Λ_L is a block diagonal matrix whose 2×2 diagonal blocks are identical to the corresponding diagonal blocks of $\widehat{\mathbf{L}}(t)$ and $\Omega_L \equiv \widehat{\mathbf{L}}(t) - \Lambda_L$ is the remaining off-diagonal part of $\widehat{\mathbf{L}}(t)$. This is a matrix version of the Jacobi iteration which is guaranteed to converge, i.e., $\widehat{\mathbf{P}}^{(k)}(t) \rightarrow \widehat{\mathbf{P}}(t)$ as $k \rightarrow \infty$, because $\widehat{\mathbf{L}}(t)$ is positive definite [42]. We initialize $\widehat{\mathbf{P}}^{(0)}(t) = \Lambda_L^{-1}$, which makes the recursion (21) equivalent (in the limit) to the series expansion used to invert a matrix in (17) [26].

The recursive scheme (21) is attractive in our suboptimal filtering scheme where $\widehat{\mathbf{L}}(t)$ is sparse at all times. Moreover, in practice typically only certain elements of $\widehat{\mathbf{P}}(t)$, namely the diagonal elements representing the variances as well as the elements near the diagonal, are desired. The recursion (21) can often be approximated effectively by updating only the diagonal and near-diagonal elements of the covariance matrix, e.g., by some nearest-neighbor or similar masking of

the matrix after each recursion [26]. In such an approximate recursion, only $O(N)$ matrix elements are updated. This, combined with the sparseness of $\widehat{\mathbf{L}}(t)$, allows practical computation of variances.

5.3 Image preprocessing

Discretization of the image sequence in time and space affects the equivalence between the intensity invariance assumption (1) and the gradient constraint (2). Dense temporal sampling of the image sequence is especially critical in practice for (2) to be useful for optical flow computation. Let the temporal sampling interval for the image sequence be Δt. Then a discrete version of (1) can be written as

$$E(\mathbf{s} + \Delta\mathbf{s}, t + \Delta t) - E(\mathbf{s}, t) = 0, \tag{22}$$

where $\mathbf{s} \equiv [s_1, s_2]^T$ and $\Delta\mathbf{s} \equiv (\Delta t)f$. By performing a Taylor series expansion of (22) we obtain

$$\begin{aligned} 0 \;=\; & \frac{\partial E}{\partial \mathbf{s}}(\Delta\mathbf{s}) + \frac{\partial E}{\partial t}(\Delta t) \\ & + \frac{\partial^2 E}{\partial \mathbf{s}\partial t}(\Delta\mathbf{s})(\Delta t) + \frac{1}{2}(\Delta\mathbf{s})^T \frac{\partial^2 E}{\partial \mathbf{s}^2}(\Delta\mathbf{s}) + \frac{(\Delta t)^2}{2}\frac{\partial^2 E}{\partial t^2} + \cdots \end{aligned} \tag{23}$$

Then, by dividing both sides by Δt, we see that (23) reduces to (2) only if all the second order partial derivatives are zero,

$$\frac{\partial^2 E}{\partial \mathbf{s}\partial t} = 0, \quad \frac{\partial^2 E}{\partial \mathbf{s}^2} = 0, \quad \text{and} \quad \frac{\partial^2 E}{\partial t^2} = 0, \tag{24}$$

or, alternatively, $\Delta t \to 0$. Thus, two ways to satisfy (2) are to increase the temporal sampling rate or to somehow reduce the high frequency components in the intensity function. The latter can be achieved by *presmoothing* or intentionally blurring the images before gradient computations [43] so that the second and higher order brightness gradients are diminished. Presmoothing also reduces the effects of noise in the brightness measurement by providing spatial averaging.

In the experiments to be presented in Sections 6, presmoothing is implemented by averaging over

9×9 local subframes[1], and improvements in accuracy (over the cases in which no presmoothing has been applied) of the optical flow estimates are observed. We further ensure the quality of the measurement by computing the second order gradients $\frac{\partial^2 E}{\partial \mathbf{s} \partial t}$, $\frac{\partial^2 E}{\partial \mathbf{s}^2}$, and $\frac{\partial^2 E}{\partial t^2}$ at each pixel after presmoothing and weighting the measurement by a function of the magnitudes of these second order gradients. Specifically, we have found experimentally (Section 6.3) that, using the weights

$$\nu(s_1, s_2, t) = \exp\left(-k \left\| \frac{\partial^2 E}{\partial \mathbf{s} \partial t} \right\|^2\right) \equiv \exp\left(-k \left\| \left[\frac{\partial^2 E}{\partial s_1 \partial t}, \frac{\partial^2 E}{\partial s_2 \partial t}\right] \right\|^2\right), \tag{25}$$

with a constant parameter k, is particularly effective in increasing accuracy of the optical flow estimates.

6 Simulations

We demonstrate the beneficial effects of the temporal coherence constraint, formulated as the dynamic model (10), for optical flow estimation by example. Synthetic image sequences of moving brightness patterns are processed by various multi-frame and single-frame optical flow estimation methods, and the improvements gained by using this particular temporal coherence constraint are discussed. The following three methods are considered:

- **SF** (Single Frame)

  This method is a discrete version of the single-frame computational approach proposed by Horn and Schunck [1]. Each frame of optical flow is computed without any provision for temporal integration of data, by solving the inversion problem (7) for $\widehat{\mathbf{f}}(t)$.

- **TCO** (Temporal Coherence, Optimally computed)

  This method is our multi-frame algorithm implemented as the *optimal* Kalman filter (11)–(16) based on the temporal coherence constraint (10).

- **TCS** (Temporal Coherence, Suboptimally computed)

  This method is the suboptimal but computationally efficient version of the **TCO** method:

[1] An alternative to this simple averaging is local fitting of an analytic surface (e.g. [44, 45]), which allows us to trade an increase in computational complexity with the advantage that image gradients can be obtained as the parameters of the surface.

The prediction step (11) of the Kalman filter is approximated as (18).

Variants of these methods arise in different computational environments. Specifically, the inversion steps (7) (for **SF**) and (16) (for **TCO** and **TCS**) can be implemented by one of the following computational procedures, leading to variations in the algorithms above:

- **dm** (direct matrix inversion)

  Although direct matrix inversion requires computational resources too large for practical optical flow estimation, it does provide us with the *exact* estimates for the algorithm comparison purposes. In one of the experiments to be presented we have chosen to use a very small image frame to perform direct inversion for such comparison purposes.

- ic (iterative inversion, iterations to convergence)

  In practice, the inversion problems are solved iteratively. We use Gauss-Seidel iterations in the experiments in this paper. Needless to say, this iterative solution will converge to the corresponding solution of **dm** in the limit.

- **is** (iterative inversion, single iteration)

  In time sequential processing, it is natural to initialize the iterative inversion at time t with the estimate obtained at time $t-1$, providing a reasonably good estimate for time t even before the first iteration. By slightly "updating" this initial guess with a single (or a small number of) Gauss-Seidel iteration(s) at the present time, a fairly accurate estimate of the flow field can emerge after continuing the process over several time-frames [1], although such estimates are suboptimal in the statistical sense.

In the experiments to be presented, each computational method is made explicit by the name of its main algorithm suffixed by the name of the variation, e.g., **TCO-dm**, **TCS-ic**, **SF-is**, etc. Also, in each experiment, the initial frame of optical flow estimate is computed identically for every participating computational method in order to highlight the differences in the temporal effects of each method. Specifically, the initial estimates are computed by either the **SF-dm** or **SF-ic** method depending on the experiment.

The method **SF-is** deserves a special attention. This method is the approach to multi-frame optical flow estimation suggested by Horn and Schunck in [1]. It performs only one Gauss-Seidel

iteration for the inverse problem (7) at each t but uses the estimate from the *previous* frame, $\widehat{\mathbf{f}}(t-1)$, to initialize the current iteration. Unlike the **SF-dm** or **SF-ic** method, therefore, this method *does* have some provision for propagating the estimates temporally. Note that if, instead of only a single Gauss-Seidel step, the iterations are allowed to converge for each frame of data, the resulting flow estimates would have lost all information from the previous frame and become exactly the same as the **SF-ic** estimates. Although the **SF-is** method is *ad hoc* in terms of its temporal integration of data, its ease in implementation is appealing from a practical point of view.

6.1 Measurement integration by temporal coherence

One of the advantages of using a temporal coherence constraint in optical flow estimation is improvement in estimates due to temporal accumulation of complementary information regarding the flow vectors. Reconstruction of optical flow using only spatial data integration (i.e., the **SF** methods) cannot be performed correctly when a complete set of the information necessary to estimate the flow vectors is not contained in each data frame. Specifically, since diversity in the orientations of the measured spatial gradients is necessary to resolve the aperture problem, optical flow computation methods employing only a spatial coherence constraint will have difficulties dealing with cases where all the spatial gradients happen to be oriented in nearly the same direction (including the cases where most of the spatial gradient vectors have small magnitudes). Addition of a temporal coherence constraint can often relieve such difficulties by allowing the use of information from adjacent image frames. We give a demonstration below that the temporal constraint is, in fact, instrumental in correctly estimating the flow in such cases.

Experiment 1: Rotating Ramp

Small image frames are used in this experiment so that the optical flow estimates of the various methods can be computed by direct matrix inversion, allowing comparison of the *exact* estimates of the different methods.

1. *The image sequence.*

   A sequence of 10×10 images of a sloping edge rotated constantly over time is observed. The amount of rotation between adjacent image frames is 0.1 radians. Figure 1 shows frames 0, 5, 10, 15, 20, and 25 of the image sequence. The sloping edge, or transition region, is a quarter-

wave of a sinusoid changing from -1 to 1 over a band approximately 5-pixel wide. The pixel values are measured with floating-point accuracy without noise. The ramp is the only region in the image with non-zero spatial gradients; the rest of the image frame is featureless (constant brightness) so that motion is undetectable there. Note that all the spatial gradient vectors in each image frame are oriented in an identical direction. The image gradients are computed as averaged first-order differences, as performed by Horn and Schunck [1].

2. *The flow estimates.*

   Figure 2 shows the estimated flow vectors using the four methods **SF-dm**, **SF-is**, **TCO-is**, and **TCO-dm** with the parameters $\rho = 1$, $\mu_1 = \mu_2 = 0.00025$, and $\mathbf{W}(t) = I$. With these values, the relative strength of spatial coherence, normalized by the strength of the brightness constraint, is about $\frac{1}{1000}$, accommodating the large spatial variations among the motion vectors in rotational motion (especially for small image frames as in this example). As described before, all four methods begin with the same initial estimates, as reflected in the results for frame 0 in the figure. The **TCO-dm** method produces a fairly accurate estimate at frame 25. The estimate by the **SF-is** method at frame 25 appears to be fairly good, also. The **SF-dm** method, however, fails completely. This behavior demonstrates that some sort of temporal integration of measurements is necessary for correct estimation in this case.

3. *The estimation errors.*

   Figure 3 displays the percent average estimation error for each t,

$$\frac{\left\|\widehat{\mathbf{f}}(t) - \mathbf{f}(t)\right\|}{\|\mathbf{f}(t)\|} \times 100, \tag{26}$$

   where $\mathbf{f}(t)$ is the true flow and $\widehat{\mathbf{f}}(t)$ is the estimated flow, for the four methods. The figure also displays the plot of standard deviation (representing the spatial variation of the estimation error $\widehat{\mathbf{f}}(t) - \mathbf{f}(t)$ for each t) associated with each of the four error curves. First, note the difference between the optimal estimates with and without the temporal coherence, as reflected by the performances of the **SF-dm** and **TCO-dm** methods whose errors are plotted as the dashed and solid curves, respectively, in the figure. Clearly, the plot for **SF-dm** displays no reduction in error as more images are processed, while the error for **TCO-dm** decreases steadily down to below 5% in the first 30 frames. Next, comparison of the error

curves for the **SF-is** (dotted line) and **SF-dm** (dashed line) methods shows that having even a weak provision for temporal data integration leads to much more accurate flow estimates than having no such provision at all. The **TCO-is** method with its *explicit* use of a temporal coherence constraint, however, performs better than the **SF-is** method, as the error curve (dash-dot line) decreases to a lower value and in a more stable (smooth and near-monotonic) manner than the curve for **SF-is**.

4. *The Kalman gains and variances for the* **TCO-dm** *method.*

   One can visualize the temporal integration process of the optimal Kalman filter in the **TCO-dm** method by observing the "images" of the magnitudes of the Kalman gains and variances. Figure 4a shows the magnitudes of the Kalman gains at frames 0, 5, 10, 15, 20, and 25. The magnitude of the Kalman gain is an indication of how much the filter values the new data in updating the estimate. Lighter pixels have higher values than darker pixels. (Note that the frame size is only 10×10, resulting in the jagged appearance of the images.) Comparison of Figures 4a and 1 reveals that the Kalman gain is high where the image contrast is high. Figure 4b shows the magnitudes of the error variances. Pixels with low (dark) variances have high confidence in their associated flow vector estimates. Notice that the area of high confidence grows with time, indicating that the filter produces good estimates of flow vectors over a wider region in the image frame as more measurements are integrated over time.

6.2 Noise reduction by temporal coherence

A temporal coherence constraint can improve the quality of optical flow estimates by reducing the effect of measurement noise through the averaging of the noisy data over time. We have added white Gaussian noise of variance 0.0025 independently to each pixel of the images in the Rotating Ramp sequence of Experiment 1. Although the magnitudes of the noise are small relative to the pixel values, the gradients computed from the corrupted images are noisy enough to make optical flow computation challenging. The sequence has been processed using the **SF-dm**, **SF-is**, **TCO-is**, and **TCO-dm** methods. Figure 5 displays the estimated flow fields, and Figure 6 shows the estimation errors (26) and their associated standard deviations for the computed optical flows. The success of the **TCO**-type methods and the failure of the **SF**-type methods are evident in the figures. The difference in the performances of the **SF-is** and **TCO-is** methods can be clearly seen in this

example (by comparing the respective 25^{th} frame estimates in Fig. 5 or the dotted and dash-dot curves in Fig. 6), signifying the explicit use of the temporal coherence constraint in the algorithm formulation.

6.3 Temporal coherence in the approximate filter

As discussed in Section 5, for realistic applications the large size of image data sets makes exact implementation of the optimal Kalman filter in the **TCO** method impractical. As a result we developed an implementable yet near-optimal filter, i.e., the **TCS** method. Here, we present numerical experiments which demonstrate the efficacy of this near-optimal **TCS** method. First we will examine the performance of the approximate filter on the small Rotating Ramp images. For these small images we can compare the output of the approximate filter to the optimal Kalman filter estimates and show that the approximate filter produces estimates that are almost indistinguishable from the optimal ones. Next we will apply our approximate filter to large images of realistic size, where the exact optimal Kalman filter cannot be used. Since the true flow field will be known, we will use the percent average estimation error (26) for each t for our flow comparisons.

We start by comparing the approximate filter with the exact, optimal Kalman filter. The noise-free image sequence of the Rotating Ramp is processed with the approximate Kalman filter, (18) (12)–(16), of the **TCS-dm** method, and the resulting estimates are compared with the estimates obtained with the corresponding optimal Kalman filter of the **TCO-dm** method. Qualitatively, the optimal and approximated optical flow estimates appear to be identical. To quantify the difference between the two estimates, we have computed for each t the difference in percent normalized average estimation errors

$$\left(\frac{\left\| \widehat{\mathbf{f}}_s(t) - \mathbf{f}(t) \right\|}{\|\mathbf{f}(t)\|} - \frac{\left\| \widehat{\mathbf{f}}_o(t) - \mathbf{f}(t) \right\|}{\|\mathbf{f}(t)\|} \right) \times 100 \tag{27}$$

where $\widehat{\mathbf{f}}_s(t)$ and $\widehat{\mathbf{f}}_o(t)$ are the estimates by the suboptimal and optimal filters, respectively, and $\mathbf{f}(t)$ is the true flow. Figure 7 shows that this approximation error is at most 3% and is negligible for most t. The error is often negative, meaning that the suboptimal filter has estimated more

accurately than the optimal filter in some frames. The figure also displays the plot of

$$\frac{\left\| \sqrt{\mathbf{p}_s(t)} - \sqrt{\mathbf{p}_o(t)} \right\|}{\left\| \sqrt{\mathbf{p}_o(t)} \right\|} \times 100 \tag{28}$$

comparing the variances (diagonals of the covariance matrices) $\mathbf{p}_s(t)$ and $\mathbf{p}_o(t)$ from the suboptimal and optimal filters, respectively. The values from the suboptimal filter are within 7% of those from the optimal filter and are within 1% for most t. A more detailed comparison may be found in [26]. The accuracy of the approximate filter along with its efficiency (both in terms of computational costs and storage requirements) allows us to impose the temporal constraint to process image sequences with a much larger and more realistic frame size than those in the experiments so far, which we do next.

Experiment 2: Stagnation Flow

In this experiment we consider estimation of the motion of a non-rigid body using the **SF-ic** and **TCS-ic** methods as well as the **SF-is** and **TCS-is** methods.

1. *The image sequence.*

   Figure 8 shows a flow pattern whose velocity vector at point (s_1, s_2) is given by $(As_1, -As_2)$ for $A = 0.1$, where the coordinate origin is at the midpoint of the bottom edge of the figure. This type of flow (for an arbitrary constant A) is useful for a local characterization of *stagnation flow* [46], i.e., the flow of fluid obstructed perpendicularly by a solid object. A sequence of 64×48 images are synthesized based on such a velocity field. Figure 9 presents four images from the sequence. Note that the direction of the predominant contrasts in each image changes from mostly vertical in the early frames to mostly horizontal in later frames, implying that some type of temporal coherence constraint is necessary for correct estimation of the flow from this image sequence. We have corrupted the images by adding an independent Gaussian noise with a variance of 9 to each pixel and then requantizing the resulting pixel values to 256 grey levels.

2. *The flow estimates and estimation errors.*

   As described in Section 5.2, the 9×9 unit uniform stencil is used to spatially smooth the

images before brightness gradients are computed. The computational parameters $\rho = 10$ and $\mu_1 = \mu_2 = 0.025$ have been used. Figure 10 shows frame 18 of the estimated flow vectors computed by the **SF-ic** and **TCS-ic** methods. The **SF-ic** method, without any provision for temporal data integration, has completely failed to estimate the flow field, while the **TCS-ic** method has performed a reasonable reproduction of the flow in Figure 8. The flows computed by the **SF-is** and **TCS-is** are shown on Figure 11, which also displays the importance of temporal coherence in estimation. The average estimation errors and associated standard deviations for the four methods, shown in Figure 12, are consistent with these observations. Again, superior performance of the **TCS**-type methods over the **SF**-type methods is displayed rather dramatically by the error curves.

3. *The number of iterations required.*

   Both the **SF-ic** and **TCS-ic** methods have been allowed to use a maximum of 500 Gauss-Seidel iterations to compute the estimates at each t; however, the actual numbers of iterations required for convergence of the solution (to within 10^{-7} rms difference from iteration to iteration) are typically lower, as shown in Figure 13. Note that both algorithms initialize each iterative session (except in the first frame) using the respective estimates from the previous frame. Figure 13 indicates that the **TCS-ic** method requires progressively fewer iterations to compute the estimates (e.g., down to 16 iterations for the estimates in frame 18) and thus has a much superior convergence property than the **SF-ic** method which requires 100 to 500 iterations for the estimates in any frame.

Experiment 3: Yosemite

Our final example involves estimation of a time-varying flow field. Ideally, the temporal dynamics of the flow field is known in advance so that a model physically more accurate than the model (10) can be constructed. Unfortunately, the dynamics of the motion vectors is seldom available prior to motion estimation in practice. In fact, the estimated optical flow vectors themselves are commonly used to infer the dynamics of the objects in the scene, which in turn determines the dynamics of the motion vectors. The dynamically simple model (10), however, can be quite effective in delivering the benefits of temporal coherence to time-varying flow fields, and this experiment describes one such case. Issues regarding modeling of time-varying flow fields are explored in more depth in [41].

1. *The image sequence.*

   A motion image sequence is synthetically generated based on a 3-D topographic model of the Yosemite Valley[2]. Figure 14 shows a typical frame from the sequence. The frame size is 316×252, and the pixel values are quantized to 256 grey levels. Figure 15 shows the tenth frame of the true optical flow sequence. As in the case with the Stagnation Flow experiment, we have corrupted the images by adding an independent Gaussian noise with a variance of 9 to each pixel.

2. *Presmoothing and gradient weighting.*

   The images are presmoothed by the 9×9 unit stencil as in the Stagnation Flow experiment. The weight for the gradients at each pixel location, i.e., each element of $\mathbf{W}(t)$, is computed using (25) with $k = \frac{1}{2}$. Compared with the uniform weighting $\mathbf{W}(t) = I$, this weighting scheme is able to decrease the average estimation error by over 4% per frame. We have attempted similar weighting schemes using other second order gradients but have not found them as beneficial as (25).

3. *The flow estimates and estimation errors.*

   The flow fields are estimated with the **SF-ic TCS-ic SF-is** and **TCS-is** methods using the parameters $\rho = 10$ and $\mu_1 = \mu_2 = 250$. We have used 500 Gauss-Seidel iterations to solve the inverse problem for each frame of flow estimate, except for the estimates in the first frame. The estimate from the previous frame is used to initialize each iterative procedure. To obtain the estimates in the first frame, due to lack of a favorable initialization values, 3500 iterations have been used. The spatial coherence applied is stronger in this example than the previous examples, in part because of sparseness of the reliable image gradients, which can be observed from the grey-scale displays of the estimation error variances on Figure 16. The dark stripes in the displays are the pixel locations where the variances are low and where the image gradients are heavily relied upon by the filter to estimate the flow field. As can be observed, the filter has taken advantage of the long and mostly linear gradients along the outlines and striations of the cliffs and mountains as well as the edges of the river. The use of a stronger spatial coherence is also justified by the mostly translational nature of the motion

[2]Lyn Quam of SRI International has produced the original image sequence.

represented by the flow field. In fact, with a larger frame size of the Yosemite image in mind, the spatial variation among the flow vectors in this example is less than that in the previous examples, allowing more rigid spatial coherence. Figure 17 shows the tenth frame of the estimated flow vectors. The noise-suppression effect of the temporal coherence constraint can be observed in the upper part of the frame. Finally, Figure 18 shows the estimation errors in the first ten frames for the four flow computation methods. The **TCS-ic** method consistently yields more accurate estimates than the other three method.

7 Conclusion

We have demonstrated that the temporal coherence afforded by the use of the dynamic model (10) can improve the quality of the optical flow estimates via temporal measurement integration and noise reduction. We have shown how to practically compute such flow estimates sequentially in time using a Kalman filter. In particular, the information form of the Kalman filter is shown to be approximable, leading to a computationally efficient formulation of an effective, general-purpose procedure for multi-frame optical flow estimation. The key to this approximation was the interpretation of the update stage of the Kalman filter as an implicitly defined, *static* spatial estimation problem for the field estimation error with a prior model specified by the current information matrix of the process. Our approximate filter then arises through the efficient specification of a *reduced-order* model of the desired sparse and banded form. Numerical experiments showed that the resulting filter provided near-optimal estimation performance. An important direction to extend this work is to consider how to deal with discontinuities in the flow field (due to object occlusion etc.) over space and time. This might involve studies on *piece-wise smooth* coherence constraints for the flow field (e.g. [32, 16]) and on spatio-temporal dynamic modeling of the discontinuity boundaries (e.g. [40]). Also, more sophisticated (and perhaps more application-specific) temporal modeling of the optical flow, such as Lagrangian modeling, is another possibility for an interesting extension to the presented work.

References

[1] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203,

1981.

[2] A. Meygret and M. Thonnat. Segmentation of optical flow and 3D data for the interpretation of mobile objects. In *Third International Conference on Computer Vision*, pages 238–245. IEEE Computer Society Press, 1990. Osaka, Japan.

[3] F. Heitz, P. Perez, E. Memin, and P. Bouthemy. Parallel visual motion analysis using multiscale Markov random fields. In *Proceedings of Workshop on Visual Motion*. IEEE Computer Society Press, 1991. Princeton, NJ.

[4] H. C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B*, 208:385–397, 1980.

[5] B. F. Buxton and H. Buxton. Monocular depth perception from optical flow by space time signal processing. *Proceedings of the Royal Society of London B*, 218:27–47, 1983.

[6] Y. Yasumoto and G. Medioni. Robust estimation of three-dimensional motion parameters from a sequence of image frames using regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:464–471, 1986.

[7] L. D. Cohen and I. Cohen. A finite element method applied to new active contour models and 3D reconstruction from cross sections. In *Third International Conference on Computer Vision*, pages 587–591. IEEE Computer Society Press, 1990. Osaka, Japan.

[8] J. L. Prince and E. R. McVeigh. Motion estimation from tagged MR image sequences. submitted to IEEE Transactions on Medical Imaging, 1991.

[9] W. J. Emery, A. C. Thomas, M. J. Collins, W. R. Crawford, and D. L. Mackas. An objective method for computing advective surface velocities from sequential infrared satellite images. *Journal of Geophysical Research*, 91:12,865–12,878, 1986.

[10] K. A. Kelly. An inverse model for near-surface velocity from infrared images. *Journal of Physical Oceanography*, 19:1845–1864, 1989.

[11] K. A. Kelly and P. T. Strub. Comparison of velocity estimates from AVHRR in the coastal transition zone. Woods Hole Oceanographic Institution, Massachusetts, 1992.

[12] A. N. Netravali and J. D. Robbins. Motion-compensated television coding: Part I. *Bell System Technical Journal*, 58:631–670, 1979.

[13] A. N. Netravali and J. D. Robbins. Motion-compensated coding: Some new results. *Bell System Technical Journal*, 59:1735–1745, 1980.

[14] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. DARPA Image Understanding Workshop*, pages 121–130, 1981.

[15] F. Glazer. Hierachical motion detection. Coins tr 87-02, University of Massachusetts, 1987. Ph.D. Thesis.

[16] H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8:565–593, 1986.

[17] S. Uras, F. Girosi, A. Verri, and V. Torre. A computational approach to motion perception. *Biological Cybernetics*, 60:79–97, 1988.

[18] P. Anandan. *Measuring Visual Motion from Image Sequences.* PhD thesis, University of Massachusetts, Amherst, 1987.

[19] D. J. Heeger. Optical flow using spatiotemporal filters. *International Journal of Computer Vision*, 1:279–302, 1988.

[20] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5:77–104, 1990.

[21] A. M. Waxman, J. Wu, and F. Bergholm. Convected activation profiles and receptive fields for real time measurement of short range visual motion. In *Proc. CVPR (IEEE), Ann Arbor*, pages 717–723, 1988.

[22] J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt. Performance of optical flow techniques. Report no. 299, Department of Computer Science, University of Western Ontario, 1991.

[23] J. J. Little and A. Verri. Analysis of differential and matching methods for optical flow. In *Proceedings of Workshop on Visual Motion*, pages 173–180. IEEE Computer Society Press, 1989. Irvine, CA.

[24] E. C. Hildreth. Computations underlying the measurement of visual motion. *Artificial Intelligence*, 23:309–354, 1984.

[25] R. Szeliski. *Baysian Modeling of Uncertainty in Low-level Vision.* Kluwer Academic Publishers, Norwell, Massachuesetts, 1989.

[26] T. M. Chin, W. C. Karl, and A. S. Willsky. Sequential filtering for multi-frame visual reconstruction. to appear in *Signal Processing*, August 1992.

[27] J. W. Woods and C. H. Radewan. Kalman filtering in two dimensions. *IEEE Transactions on Information Theorey*, IT-23:473–482, 1977. (also see "Correction" in IT-25: 628-629).

[28] M. S. Murphy and L. M. Silverman. Image model representation and line-by-line recursive restoration. *IEEE Transactions on Automatic Control*, AC-23:809–816, 1978.

[29] J. W. Woods. Two-dimensional Kalman filtering. In T. S. Huang, editor, *Two-Dimensional Digital Signal Processing I*, pages 155–205. Springer-Verlag, 1981.

[30] M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76:869–889, 1988.

[31] N. M. Grzywacz, J. A. Smith, and A. L. Yuille. A common theoretical framework for visual motion's spatial and temporal coherence. In *Proceedings of Workshop on Visual Motion*, pages 148–155. IEEE Computer Society Press, 1989. Irvine, CA.

[32] M. J. Black and P. Anandan. A model for the detection of motion over time. In *Third International Conference on Computer Vision*, pages 33–37. IEEE Computer Society Press, 1990. Osaka, Japan.

[33] A. Singh. Incremental estimation of image-flow using a Kalman filter. In *Proceedings of Workshop on Visual Motion*, pages 36–43. IEEE Computer Society Press, 1991. Princeton, NJ.

[34] B. D. O. Anderson and J. B. Moore. *Optimal Filtering.* Prentice-Hall, Englewood Cliffs, N.J., 1979.

[35] A. Gelb, editor. *Applied Optimal Estimation.* MIT Press, Cambridge, MA, 1974.

[36] F. L. Lewis. *Optimal Estimation.* John Wiley & Sons, New York, 1986.

[37] B. C. Levy, M. B. Adams, and A. S. Willsky. Solution and linear estimation of 2-D nearest-neighbor models. *Proceedings of IEEE*, 78:627–641, 1990.

[38] D. Terzopoulos. Image analysis using multigrid relaxation models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8:129–139, 1986.

[39] A. Rougee, B. C. Levy, and A. S. Willsky. An estimation-based approach to the reconstruction of optical flow. Technical Report LIDS-P-1663, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 1987.

[40] T. M. Chin and A. J. Mariano. Optimal space-time interpolation of gappy frontal position data. In *'Aha Huliko'a Hawaiian Winter Workshop (Probability Concepts in Physical Oceanography)*, 1993.

[41] T. M. Chin. *Dynamic Estimation in Computational Vision.* PhD thesis, Massachusetts Institute of Technology, 1991.

[42] G. H. Golub and C. F. van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, Maryland, 1989.

[43] J. K. Kearney, W. B. Thompson, and D. L. Boley. Optical flow estimation: An error analysis of gradient-based methods with local optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9:229–244, 1987.

[44] D. M. Martinez. *Model-Based Motion Estimation and its Application to Restoration and Interpolation of Motion Pictures.* PhD thesis, Massachusetts Institute of Technology, 1986.

[45] M. K. Brown and B. Shahraray. Robust depth estimation from optical flow. In *Proc. Computer Vision and Pattern Recognition (Tampa, Florida)*, pages 641–650, 1988.

[46] M. C. Potter and J. F. Foss. *Fluid Mechanics.* Great Lakes Press, Okemos, Michigan, 1982.

List of Figures

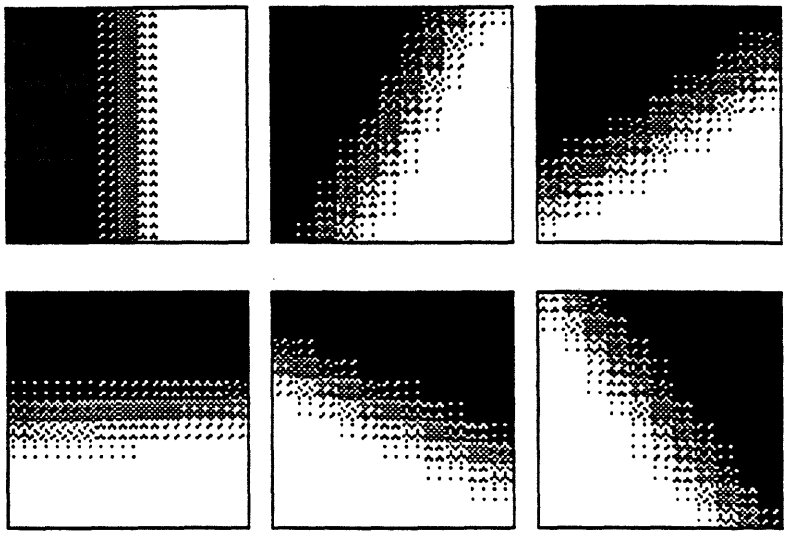


Figure 1: The Rotating Ramp. Frames 0, 5, 10, 15, 20, and 25 are shown.

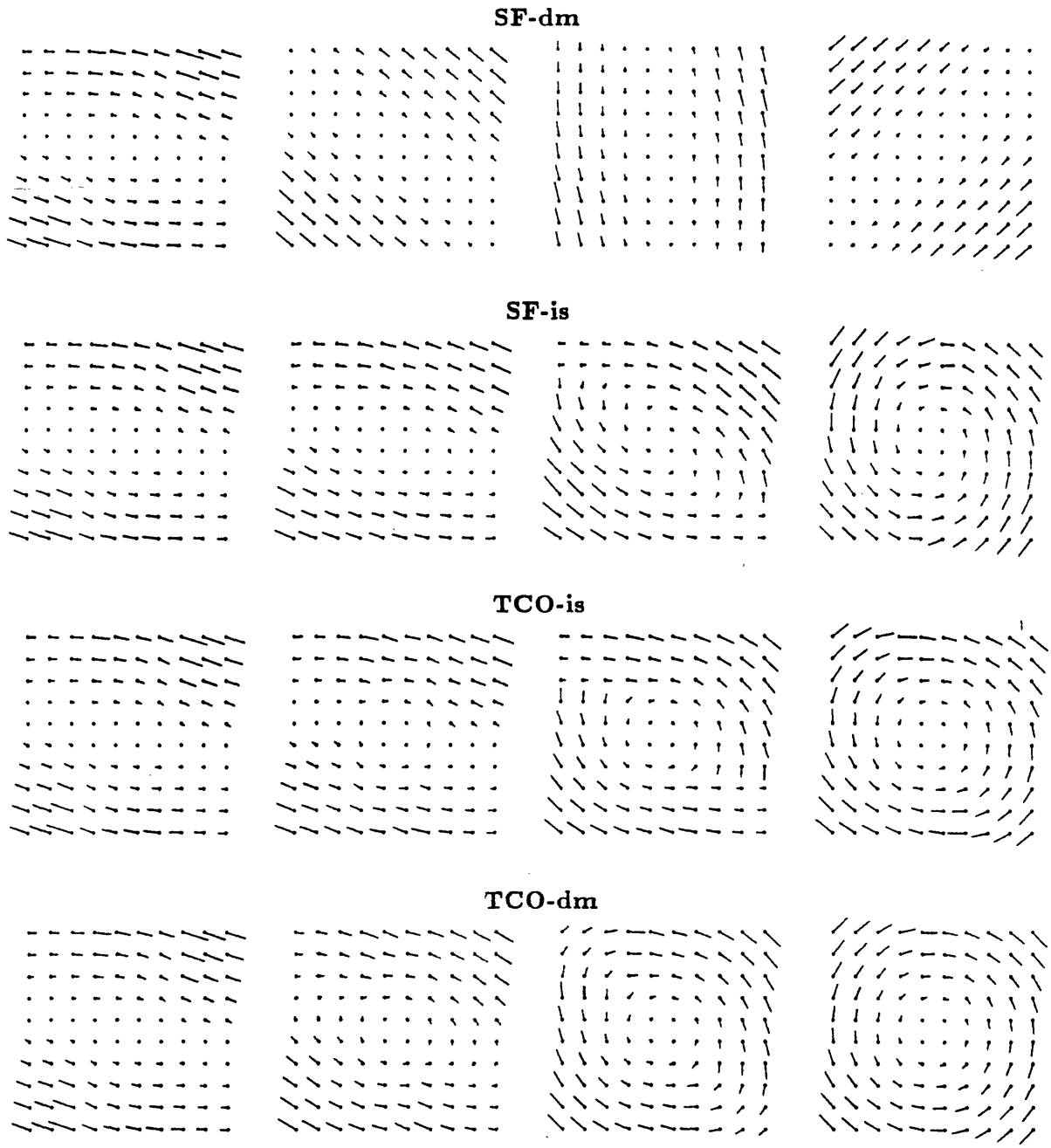
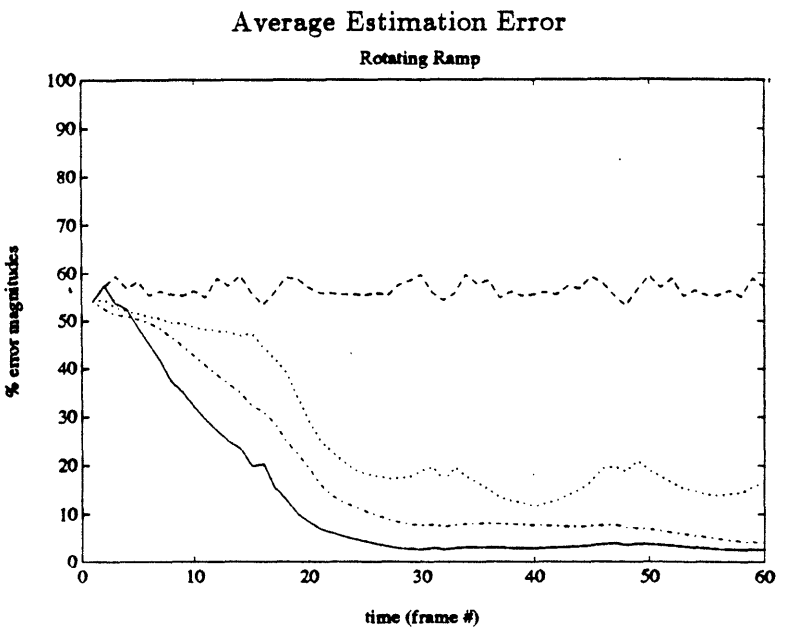


Figure 2: Optical flow estimates for the Rotating Ramp example. The flow patterns at frames (from left to right) 0, 5, 15, and 25 are shown. The flow vectors are magnified by 1.5 for clarity.



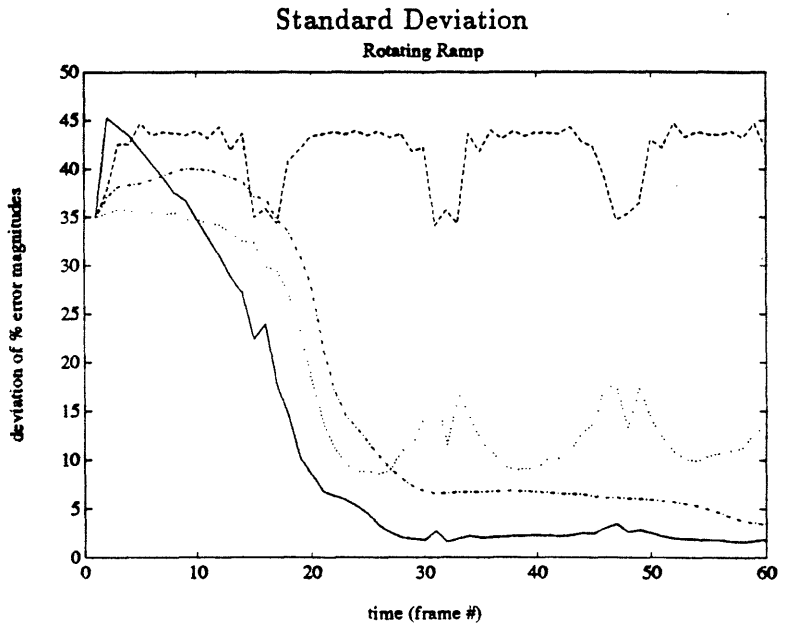


Figure 3: The average estimation errors and their associated standard deviations in the Rotating Ramp example for the four methods: **SF-dm** (dashed line), **SF-is** (dotted line), **TCO-is** (dash-dot line), and **TCO-dm** (solid line).

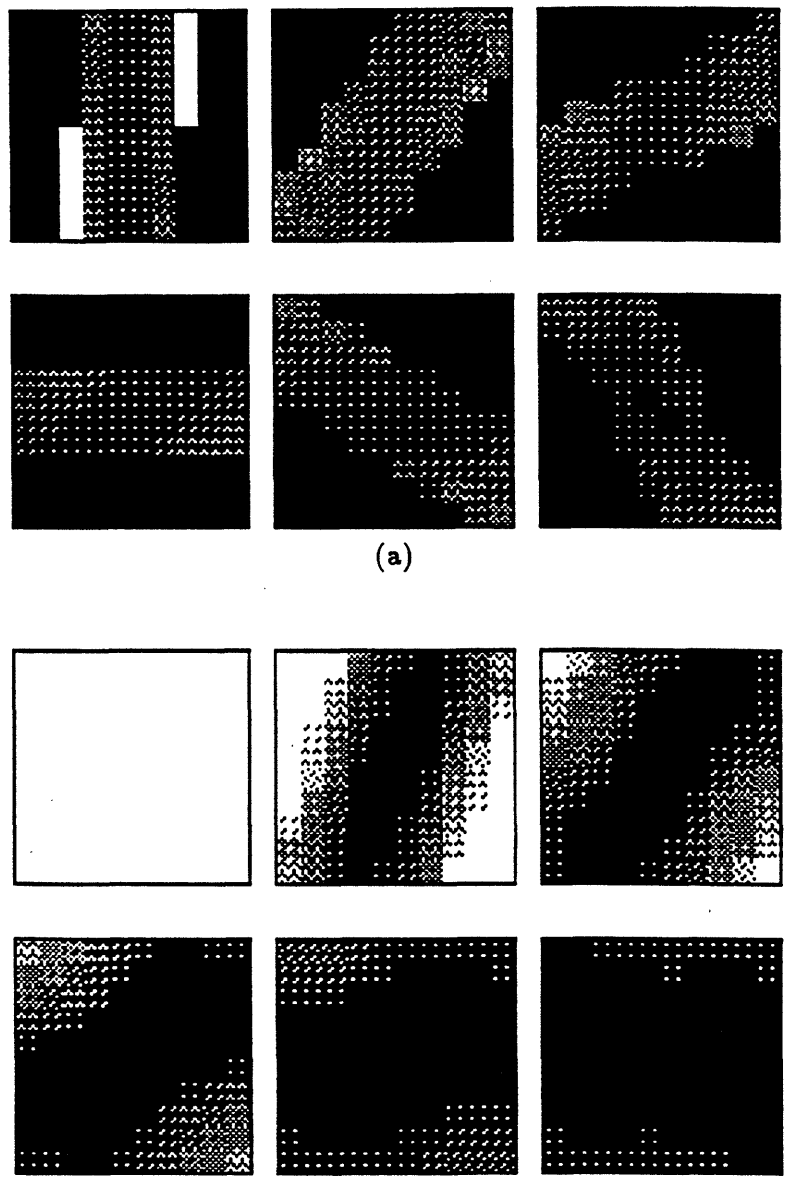


Figure 4: The magnitudes of (a) the Kalman gain and (b) error variance for the **TCO-dm** method in the Rotating Ramp example. Frames 0, 5, 10, 15, 20, and 25 are shown.

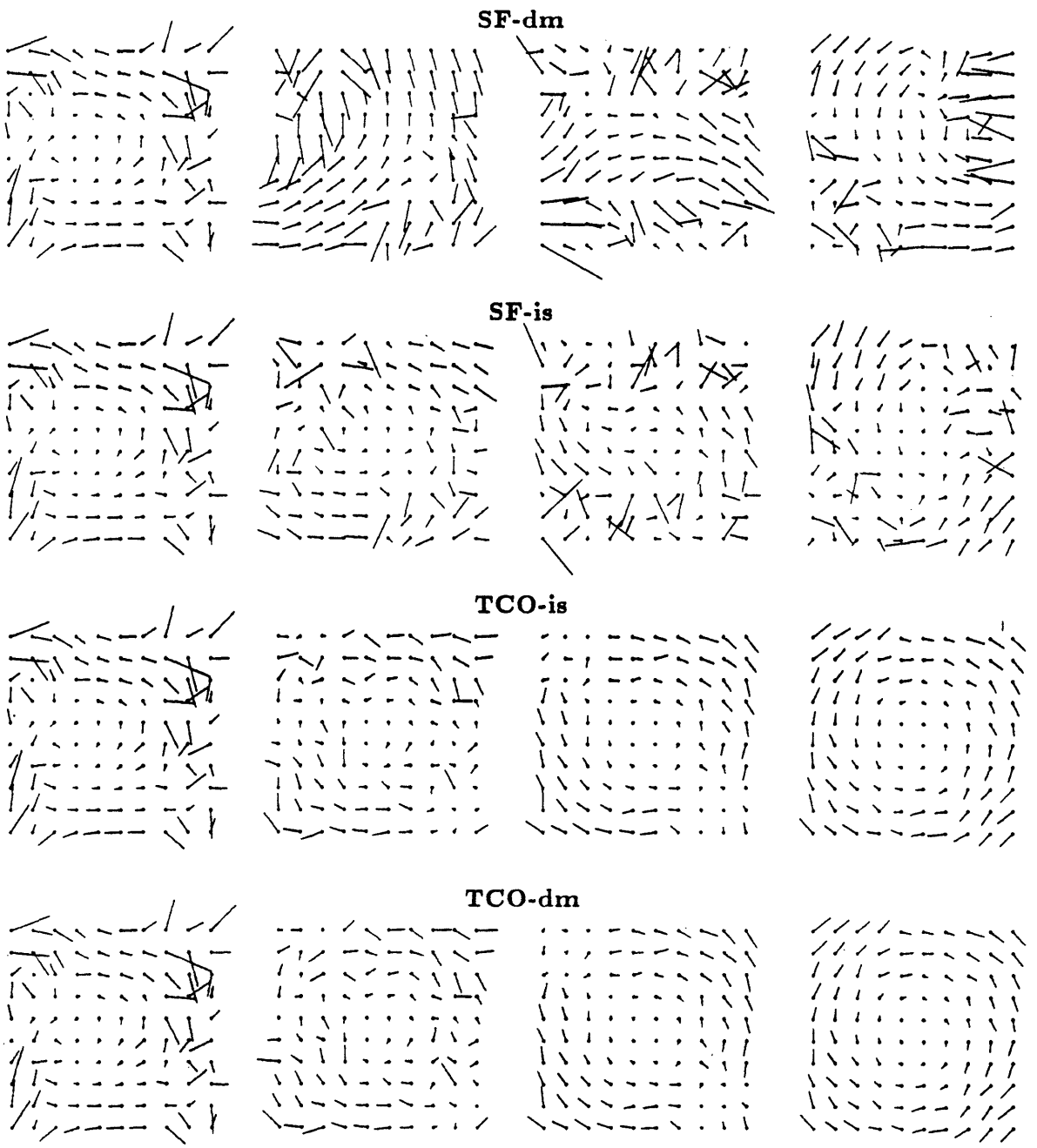
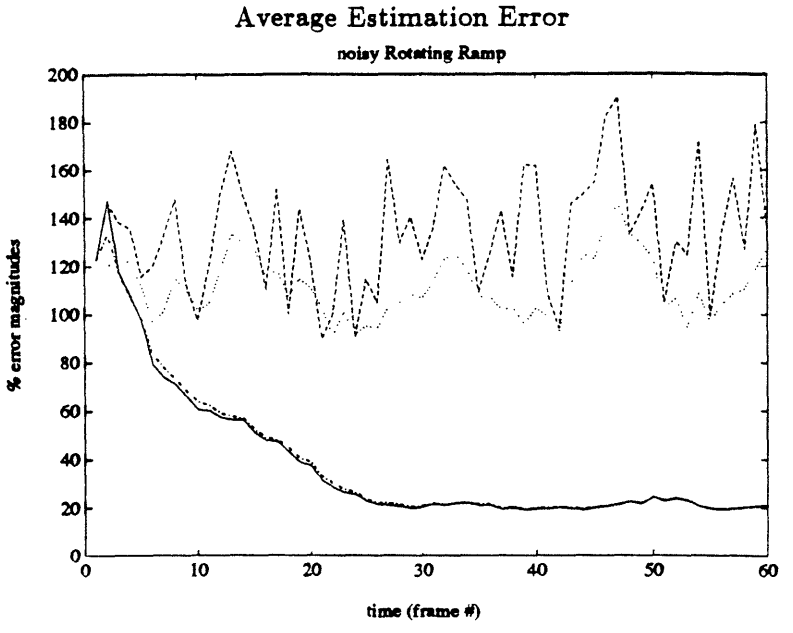


Figure 5: Optical flow estimates for the *noisy* Rotating Ramp example. The flow patterns at frames (from left to right) 0, 5, 15, and 25 are shown. The flow vectors are magnified by 1.5 for clarity.



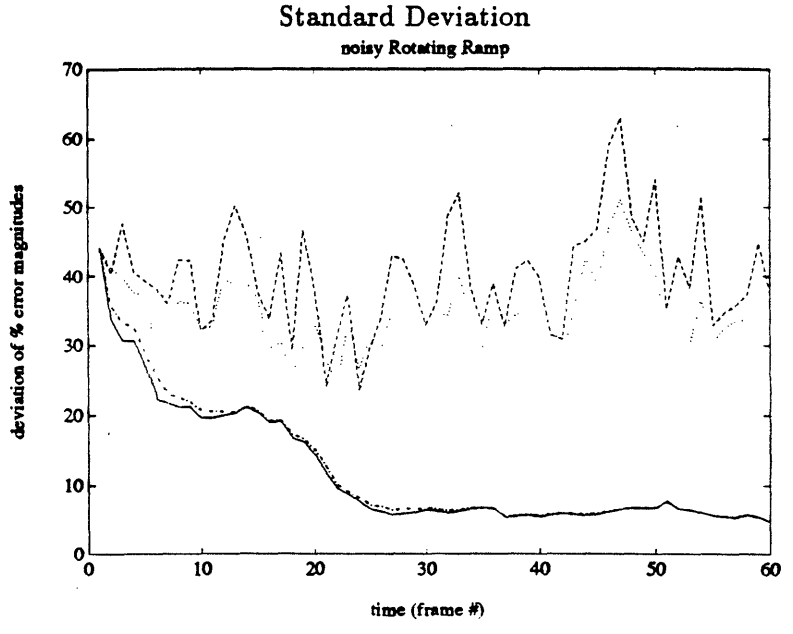


Figure 6: The estimation errors and associated standard deviations when the *noisy* Rotating Ramp image sequence is processed by the three methods: **SF-dm** (dashed-line), **SF-is** (dotted line), **TCO-is** (dash-dot line) and **TCO-dm** (solid line).

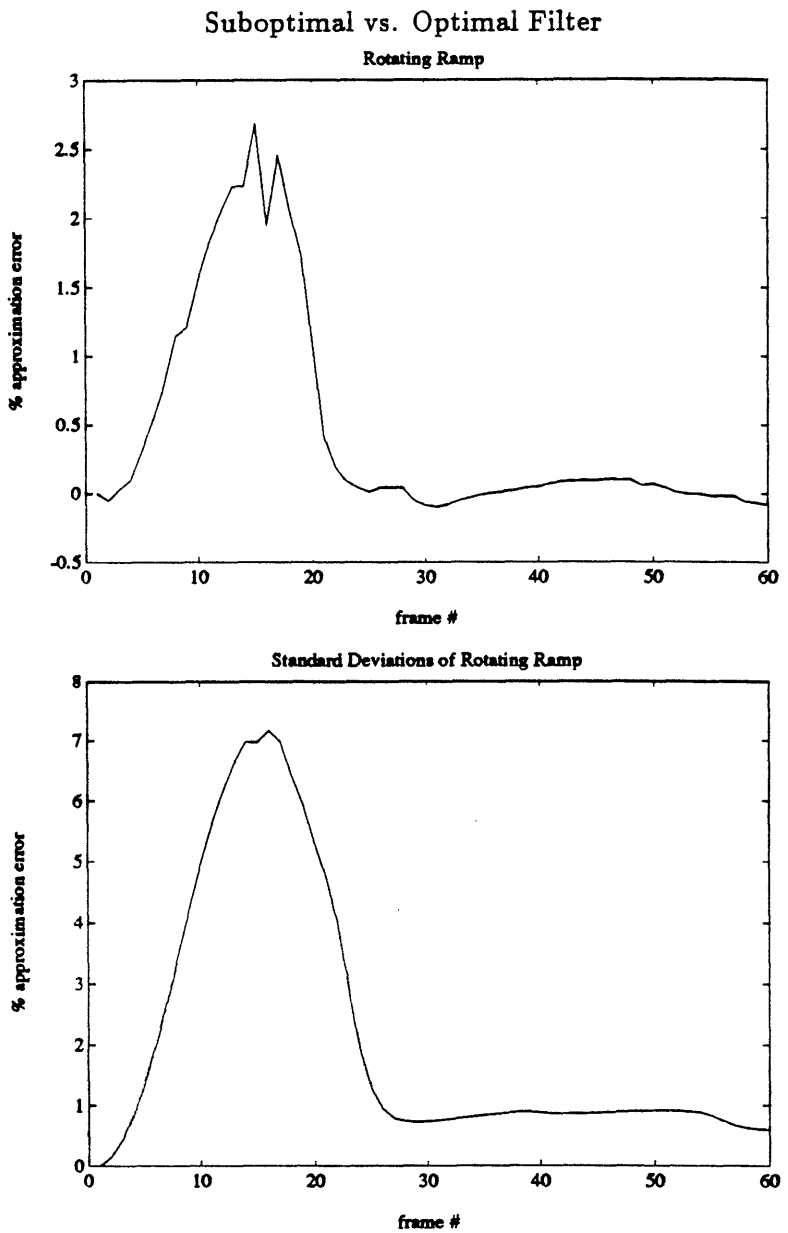


Figure 7: The error introduced when the optimal Kalman filter in the **TCO-dm** method is approximated as in the **TCS-dm** method.

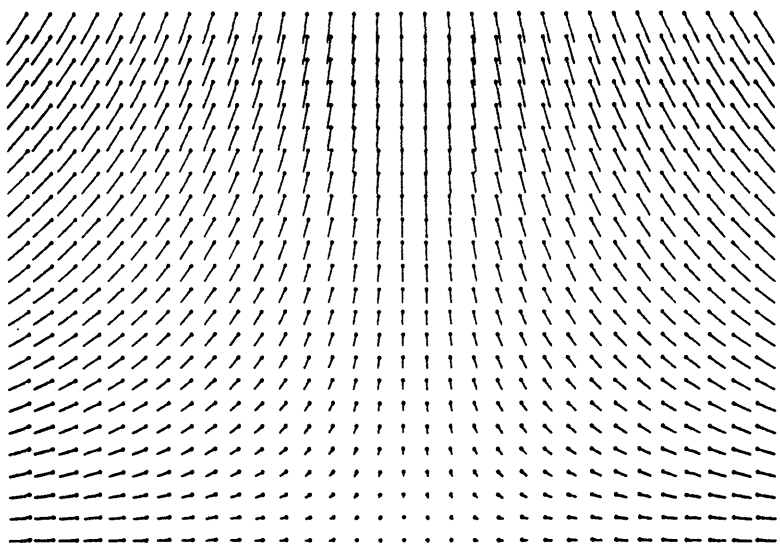


Figure 8: The true flow in the Stagnation Flow experiment. Every other flow vector along each axes is shown with a magnification factor of 4 for clarity.

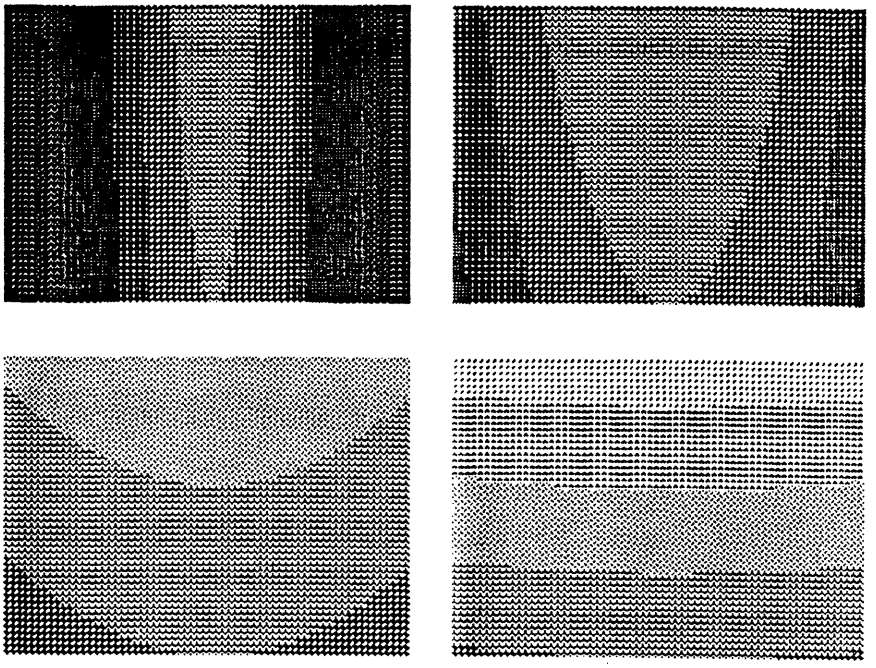


Figure 9: The Stagnation Flow image sequence. Frames 0 and 7 (top row) as well as 14 and 21 (bottom row) are shown.

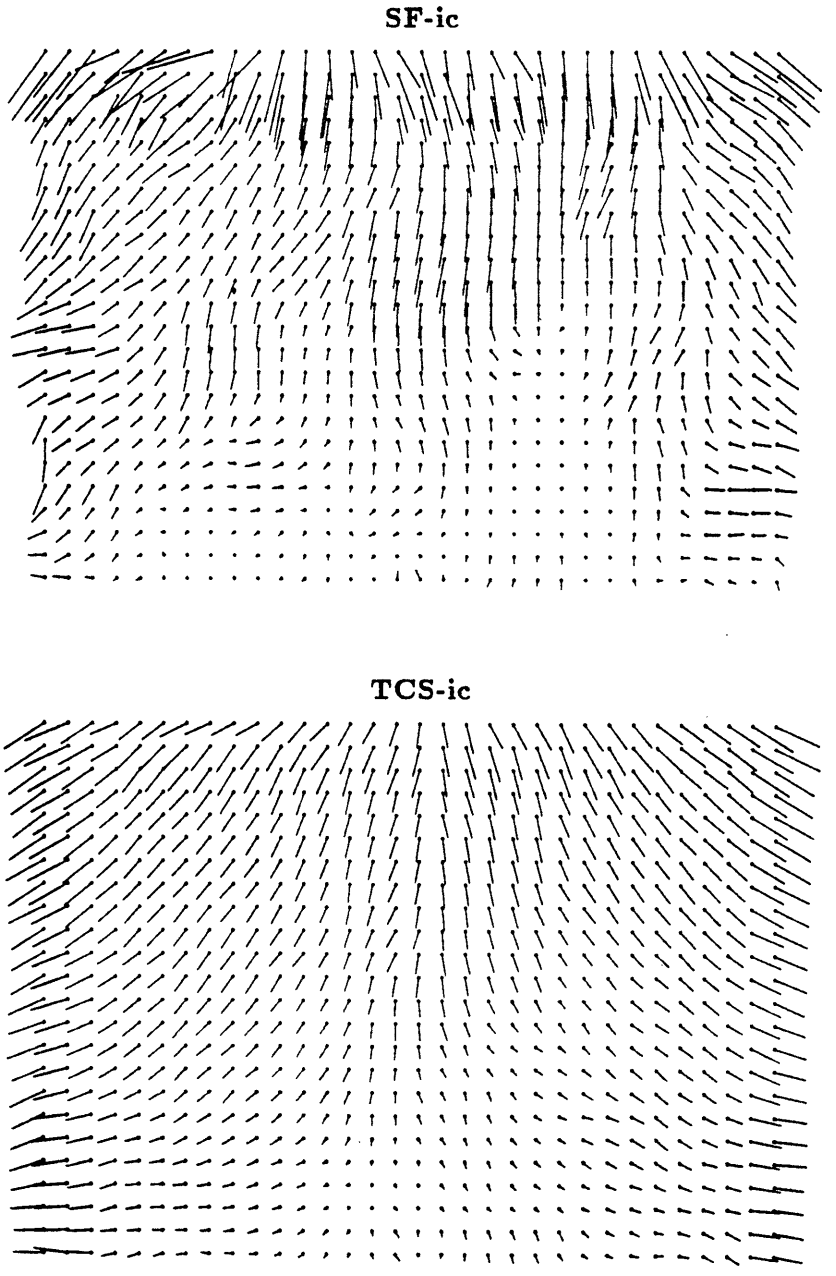


Figure 10: The optical flow estimates for frame 18 of the Stagnation Flow sequence by the **SF-ic** and **TCS-ic** methods.

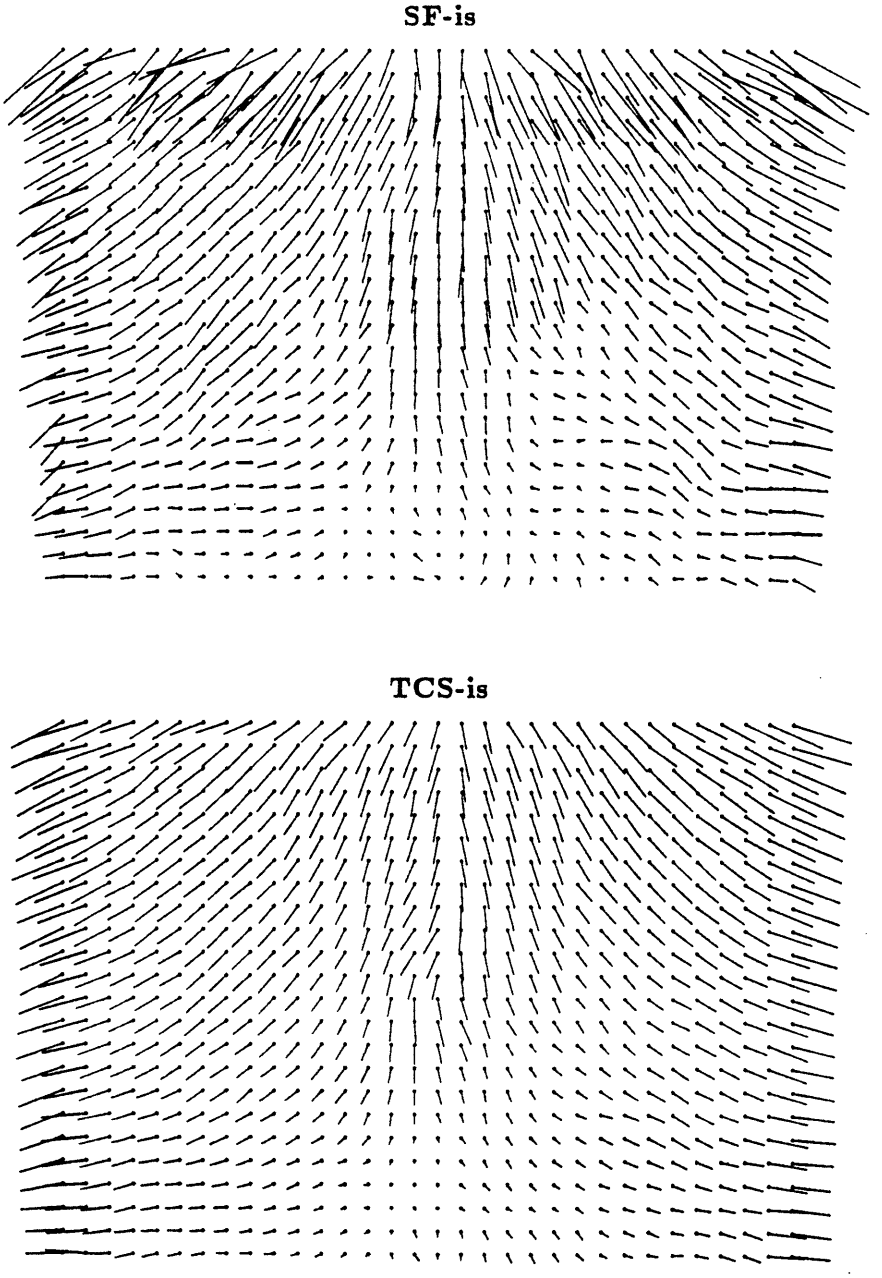
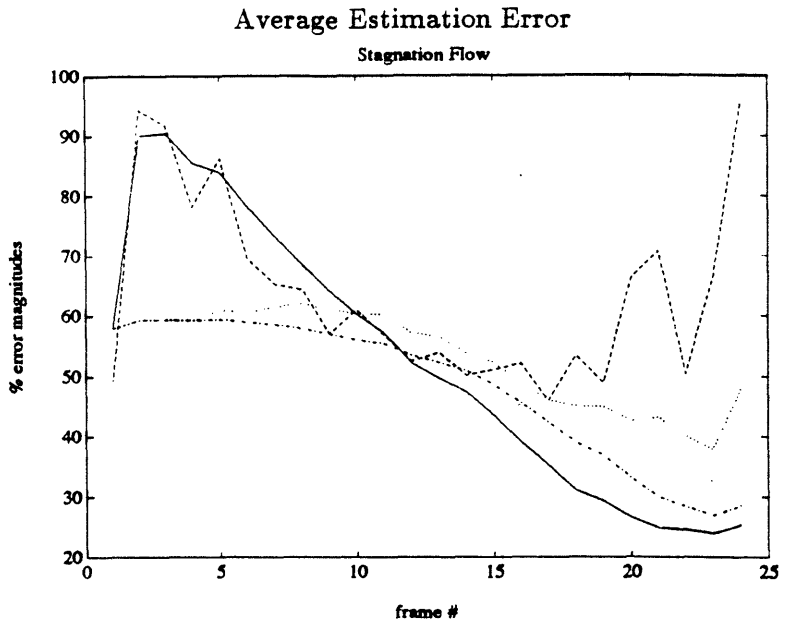


Figure 11: The optical flow estimates for the frame 18 of the Stagnation Flow sequence by the **SF-is** method and **TCS-is** method.



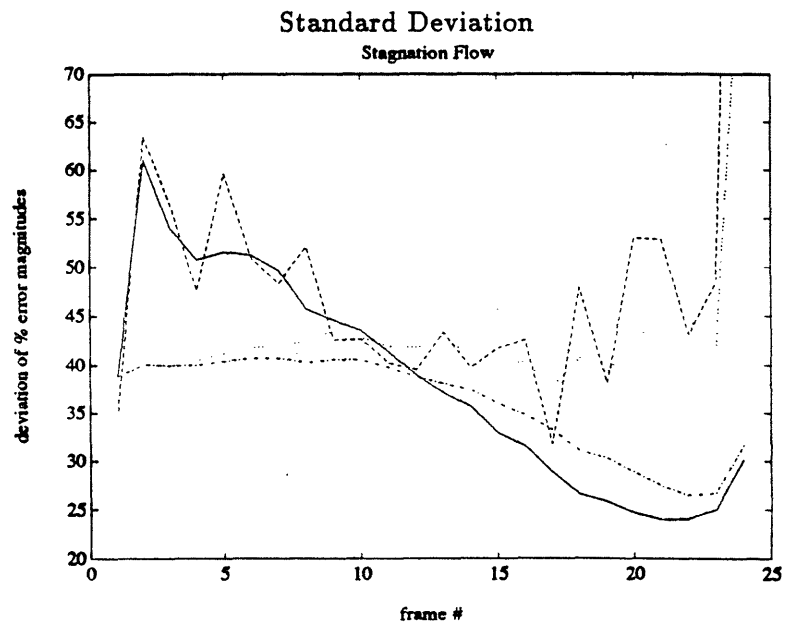


Figure 12: The estimation errors and associated standard deviations by the **TCS-ic** (solid-line), **SF-ic** (dashed-line), **SF-is** (dotted-line), and **TCS-is** (dash-dot line) methods for the Stagnation Flow experiment.

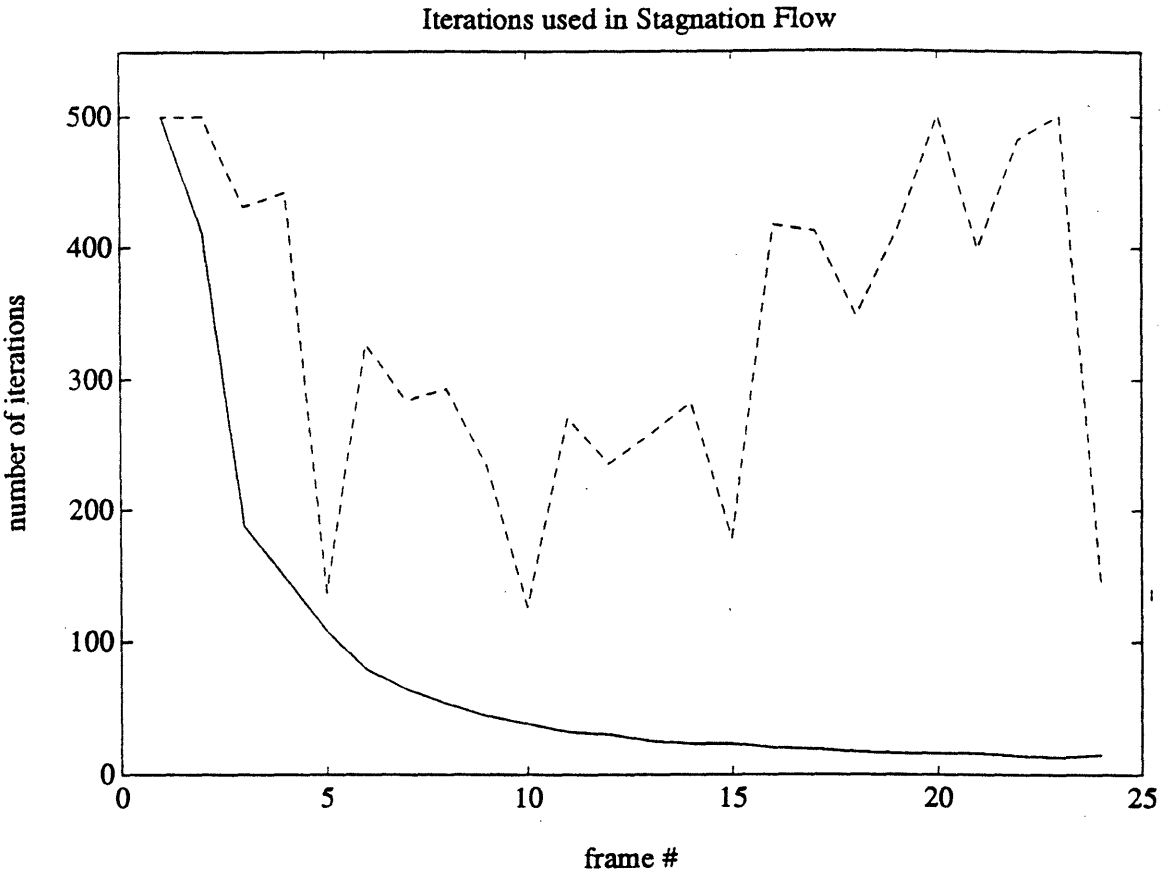


Figure 13: The number of iterations used by the **TCS-ic** (solid-line) and **SF-ic** (dashed-line) methods for convergence of the estimates in the Stagnation Flow experiment.

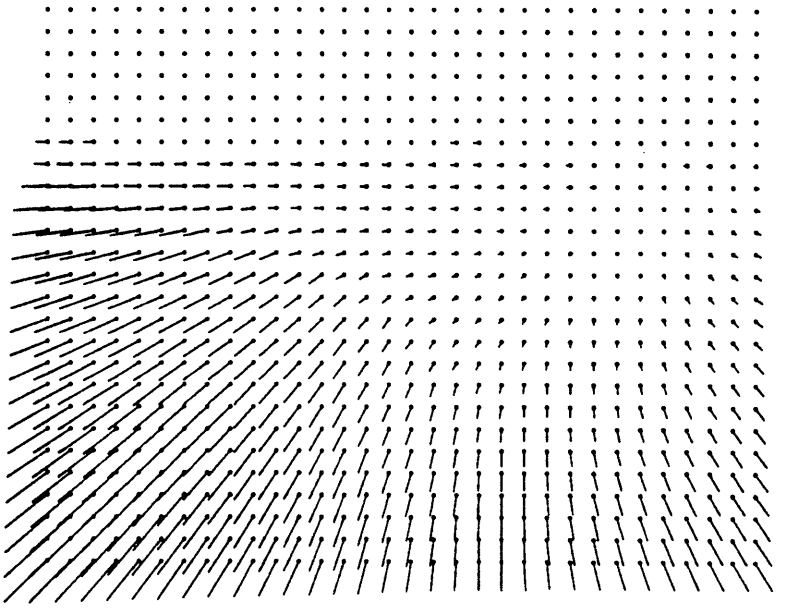


Figure 15: The true flow for the tenth frame in the Yosemite image sequence. The flow vectors are reduced by a factor of 2 for display purposes.

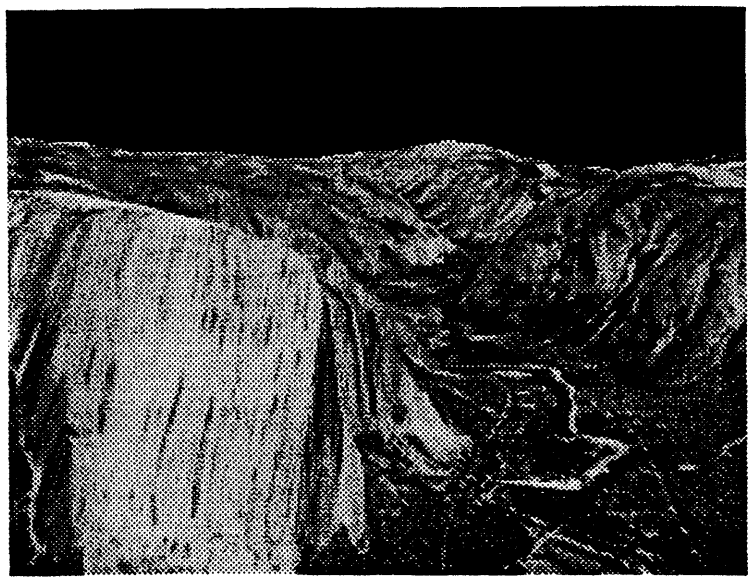
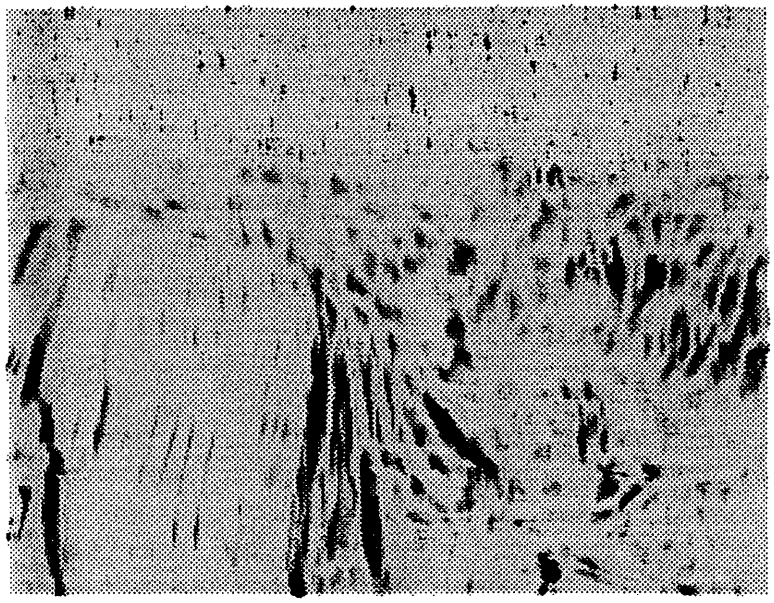


Figure 14: A frame from the Yosemite image sequence

Variances in the Horizontal Components



Variances in the Vertical Components

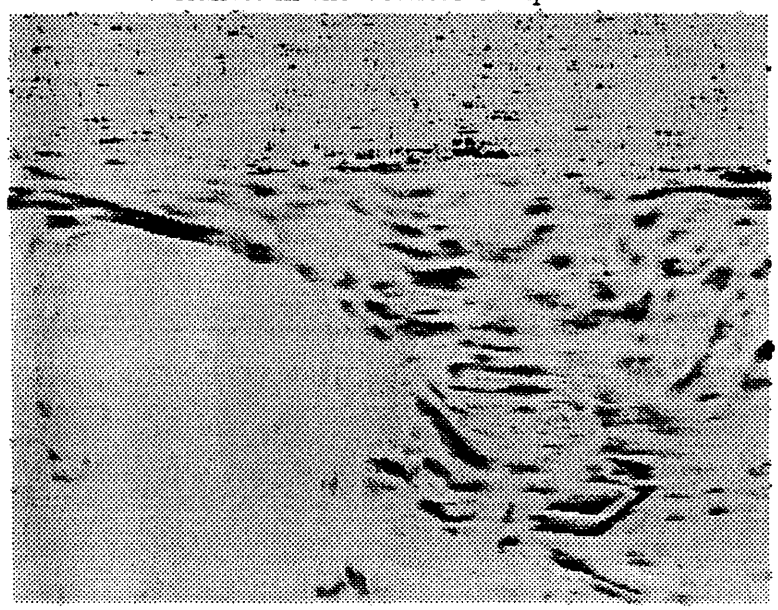


Figure 16: The filter variances associated with the two components of each flow vector estimate for the second frame in the Yosemite image sequence.

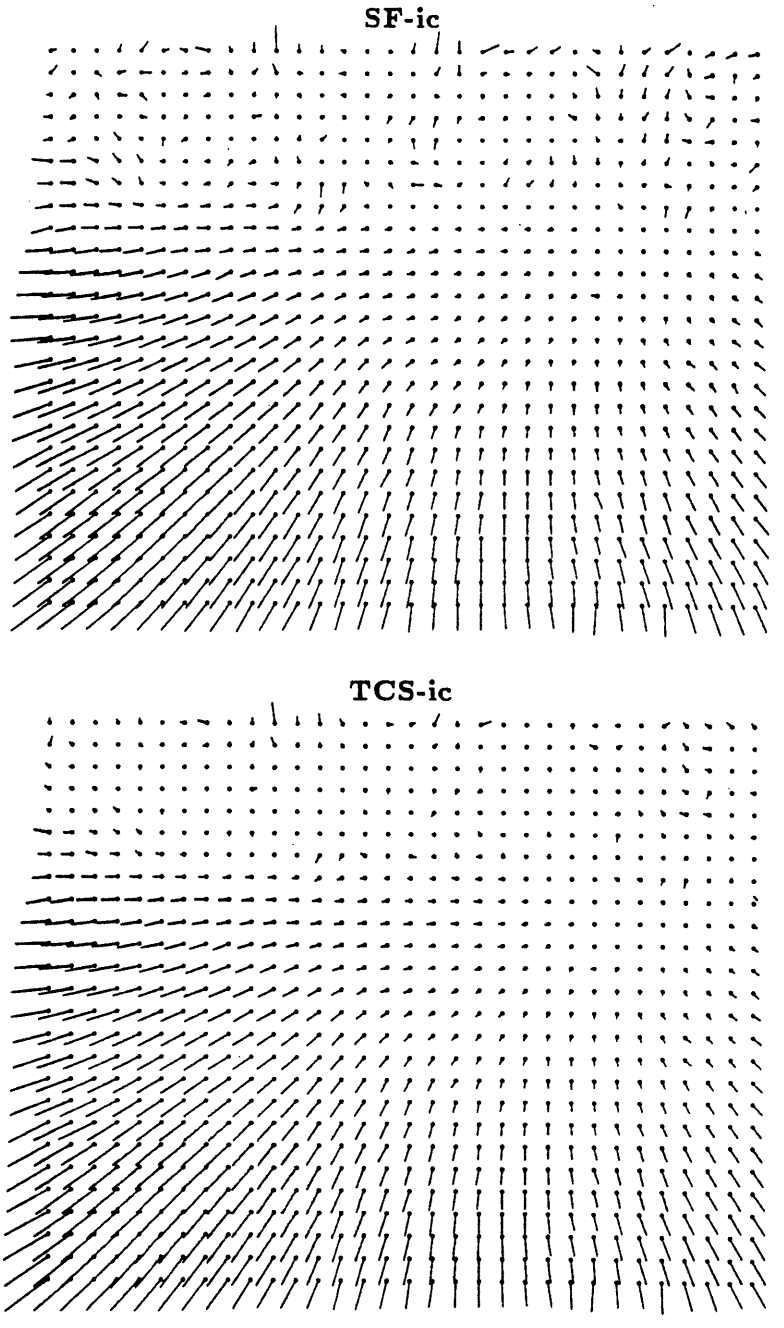


Figure 17: The optical flow estimates for the tenth image frame by the **SF-ic** and **TCS-ic** methods. The flow vectors are reduced by a factor of 2 for display purposes.

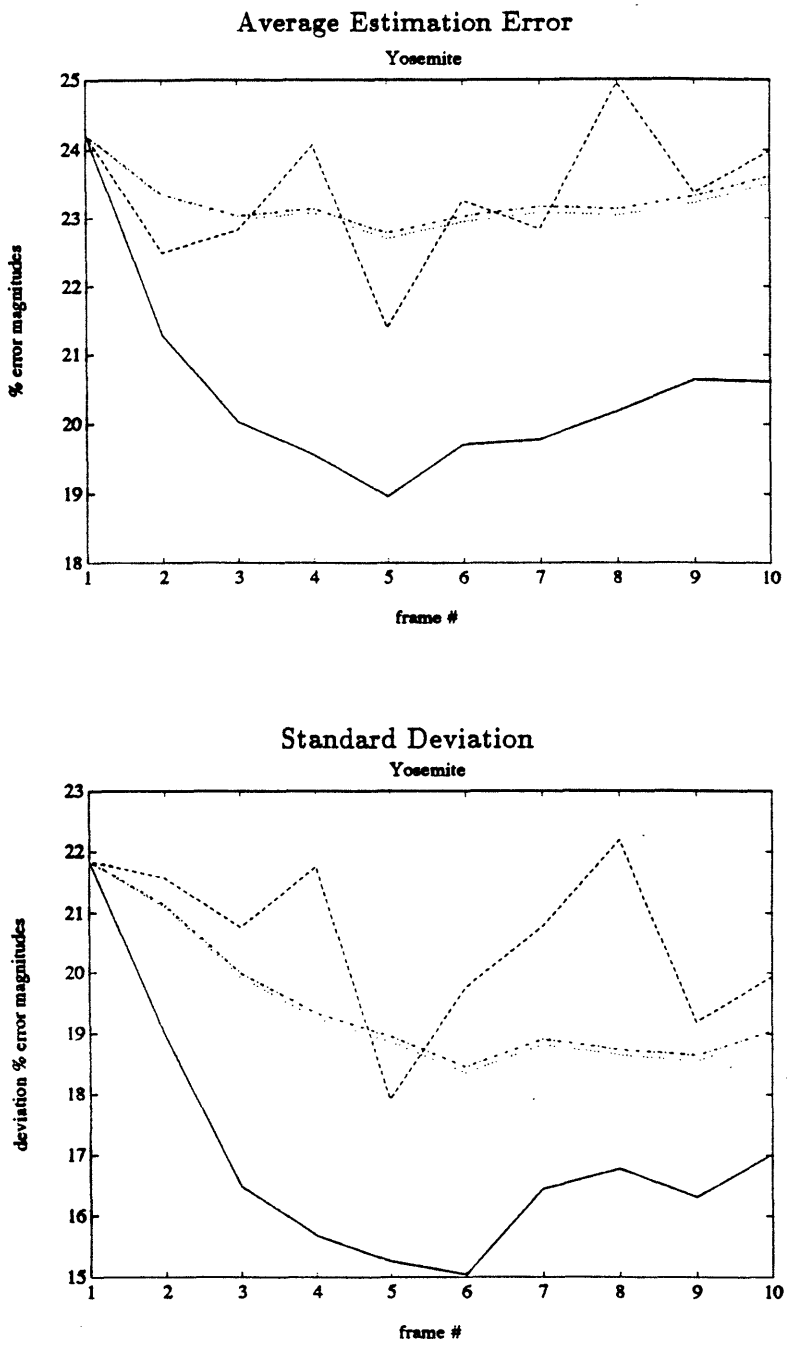


Figure 18: The estimation errors and associated standard deviations by the **TCS-ic** (solid-line), **SF-ic** (dashed-line), **SF-is** (dotted-line), and **TCS-is** (dash-dot line) methods when the noisy Yosemite image sequence is processed.