Probabilistic and Sequential Computation of Optical Flow Using Temporal Coherence

Toshio M. Chin, *Student Member, IEEE,* William C. Karl, and Alan S. Willsky, *Fellow, IEEE*

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, multiframe 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 multiframe optical flow estimation problem, providing a solution that is sequential and recursive in time. Experiments are presented to demonstrate that the resulting multiframe estimates are more robust to noise than those provided by the original, single-frame formulation. In addition, 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.**

I. 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 hierarchical 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]–[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 oceanic and atmospheric processes [9], [10]. 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 [11], [12]. There exist a variety of techniques for computing optical flow, including [1], [13]–[20], as well as comparative studies of them [21], [22]. 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, which is 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 [23]), 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 descriptions of the unknown field [24], [25]. Specifically, the optical flow formulation by Horn and Schunck can be considered to be a Bayesian estimation problem with additive Gaussian noise. Utilizing such an

Manuscript received June 9, 1992; revised July 13, 1993. This work 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 DAAL03-92-G-0115. The associate editor coordinating the review of this paper and approving it for publication was Dr. Homer H. Chen.

T. M. Chin is with the Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL 33149.

W. C. Karl is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA USA.

A. S. Willsky is with the Laboratory for Information and Decision Systems and the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA USA.

IEEE Log Number 9404449.

estimation-theoretic framework for optical flow computation, we model the time-varying unknown flow field as a dynamic process with an associated evolution equation that captures the temporal coherence constraint of the variational formulation. Thus, the temporal extension of the Horn and Schunck formulation, i.e., multiframe optical flow 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], [24]. The probabilistic framework allows quantification of uncertainty in the estimates through computation of the second-order statistics.

Although Kalman filtering allows time-recursive estimation of multiframe 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, which are 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 that 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]–[29]; the computational algorithm in this paper represents a 3-D version for image sequences in which we must determine a reduced-order spatial 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 II, we review the classical single-frame optical flow estimation problem in a continuous setting. In Section III, we present our temporal extension to the continuous classical problem. In Section IV, 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 multiframe problem. In Section V, we investigate implementation issues, including approximation of the Kalman filter and the effects of discretization on the fundamental measurement constraint. In Section VI, we present various simulation results demonstrating the benefits of applying temporal coherence to multiframe 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 VII. Preliminary results of parts of this work have appeared in [25].

II. Optical Flow Estimation

We perceive motion by temporally tracking image intensity patterns that are 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 $f(s_1, s_2, t) \equiv [ds_1/dt, ds_2/dt]^T$ 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 [23] 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 the solution of a quadratic minimization problem

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

where $\nu(s_1, s_2, t) \neq 0$, and 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 [23].

III. Multiframe Formulation

We now consider imposition of *temporal coherence* [31] to the flow field in addition to the more commonly used spatial coherence enforced by (3), thus allowing the utilization of more data (the gradients of the image intensity) for each frame of flow vector estimates. A temporal coherence imposes an inertia condition on the flow field, favoring smooth changes in the optical flow vectors over time. Models of optical flow incorporating temporal coherence are applicable to a wide range of motions in natural scenes, as most motions display inertia of some type. A simple temporal extension [24], [26] of (3) is used to obtain such a *multiframe* formulation of the optical flow computation problem. In particular, for $0 \leq t \leq \tau$, we find the flow field $\hat{f}(s_1, s_2, t)$, which provides the solution

to

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

Note that (4) is obtained by the addition to (3) of a quadratic term involving the first-order temporal derivative of f. The inclusion of the temporal constraint allows the integration of data over time. As demonstrated in Section VI, 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 nontemporal formulation (3).

Typically, it is desirable to compute an optical flow field corresponding to each new image frame as soon as the image data are recorded. In terms of the multiframe formulation (4), this means that the *most recent* (i.e., $t = \tau$) solution $\hat{f}(s_1, s_2, \tau)$ must be computed via optimization of a *distinct* 3-D optimization problem for each τ as τ increases. It is a seemingly prohibitive computational task, particularly when the additional need for calculation of the uncertainty in the solution is also taken into account. Such solutions, however, can be computed efficiently by a Kalman filter [32]–[34], which has an attractive time-sequential computational structure wherein the flow estimate at the current time is recursively *updated* based on the new data, thus allowing us to calculate $\hat{f}(s_1, s_2, \tau)$ without resorting to repeated 3-D optimizations.

Despite their efficiency, Kalman filters as applied to image data still represent computationally intensive tasks. In previously reported applications of Kalman filtering algorithms to optical flow estimation [35], [36] (as well as to other "low-level" computational vision problems [24]), the formulations are simplified apparently to reduce such computational complexity. 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 [32]–[34] and allows for the optimal fusing of the current estimate with new observations. In this paper, we employ a more systematic and rational approach to address this computational issue and derive a computationally efficient yet near-optimal approximation to the Kalman filter algorithm for the multiframe optical flow estimation problem. The mathematical details of our approximation techniques can be found in [26] in the more general context of low-level visual reconstruction.

IV. Discretization and Probabilistic Interpretation

A. 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 $\boldsymbol{f}(t)$ be a vector or 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, $\boldsymbol{f}(t)$ has $2N$ elements. We similarly define $\boldsymbol{g}(t)$ to be an N-vector of the lexicographically ordered samples of $(-\partial E/\partial t)$ taken on the same grid. Finally, let $\boldsymbol{H}(t)$ and $\boldsymbol{W}(t)$ be the block diagonal matrices whose diagonal elements are the samples of $[\partial E/\partial s_1, \partial E/\partial s_2]$ 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

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

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

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

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

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

Solving the quadratic minimization problem (5) is equivalent to solving a maximum likelihood estimation problem [34] for $\boldsymbol{f}(t)$ with the following *observation equation:*

$$\begin{bmatrix} \boldsymbol{g}(t) \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}(t) \\ \boldsymbol{S}_1 \\ \boldsymbol{S}_2 \end{bmatrix} \boldsymbol{f}(t) + \boldsymbol{r}(t),$$

$$\boldsymbol{r}(t) \sim \left(0, \begin{bmatrix} \boldsymbol{W}^{-1}(t) & & \\ & \mu_1^{-1} I_{2(N-n_2)} & \\ & & \mu_2^{-1} I_{2(N-n_2)} \end{bmatrix} \right) \quad (6)$$

where we have used the notation $\boldsymbol{x} \sim (\boldsymbol{m}, \boldsymbol{C})$ to denote a Gaussian random vector $\boldsymbol{x}$ whose mean and covariance are $\boldsymbol{m}$ and $\boldsymbol{C}$, respectively. Thus, $\boldsymbol{r}(t)$ is a zero-mean Gaussian random noise process. Note that $\boldsymbol{W}^{-1}(t)$ is a diagonal covariance matrix, whose nonzero 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). In addition, μ_1^{-1} and μ_2^{-1} are 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 $\hat{\boldsymbol{f}}(t)$ is obtained as the solution of the inverse problem

$$\boldsymbol{L}(t)\hat{\boldsymbol{f}}(t) = \boldsymbol{H}^T(t)\boldsymbol{W}(t)\boldsymbol{g}(t) \quad (7)$$

where $\boldsymbol{L}(t) = \boldsymbol{H}^T(t)\boldsymbol{W}(t)\boldsymbol{H}(t) + \mu_1\boldsymbol{S}_1^T\boldsymbol{S}_1 + \mu_2\boldsymbol{S}_2^T\boldsymbol{S}_2$. Equation (7) specifies a discrete version of the coupled Poisson equations of the Horn and Schunck formulation. The matrix operator $\boldsymbol{L}(t)$ has a sparse, *nearest neighbor* (a nested block tridiagonal) structure [37], whose sparseness enables us to use efficient iterative procedures, such as multigrid methods [38], in the solution of (7).

The matrix $\boldsymbol{L}(t)$ is the *information matrix* (the inverse of the covariance matrix) associated with the posterior estimation error $\boldsymbol{e}(t) \equiv \boldsymbol{f}(t) - \hat{\boldsymbol{f}}(t)$, i.e., $\boldsymbol{e}(t) \sim (0, \boldsymbol{L}^{-1}(t))$. It is insightful to interpret this information matrix as a Markov random field (MRF) model specification for the estimation error process [24], [26]

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

The nearest-neighbor structure of the matrix $\boldsymbol{L}(t)$ reflects the *neighborhood* (or the extent of local interactions among the components of $\boldsymbol{e}(t)$) in the MRF model [39]. The size of the neighborhood determines the *order* of an MRF model, just as the number of parameters determines the order of an autoregressive model. As we have just noted, the small size of the neighborhood facilitates the computation of the estimate by inversion of (7). An advantage of the MRF modeling framework is that it allows modeling/estimation of the pixel process as well as "line process" (e.g., contours traced by discontinuities in the pixel process) simultaneously [40]. An MRF-based motion estimation scheme can thus be expanded into an algorithm capable of coestimating the motion discontinuities (e.g., due to object boundaries), as demonstrated in [35].

B. Multiframe Processing

The continuous optimization problem (4) for multiframe optical flow computation can be considered to be an *optimal smoothing* problem based on the temporal Gauss–Markov dynamic equation $\partial/\partial t\, f(s_1, s_2, t) = q(t)$, where $q(t)$ is a Gaussian white noise process of zero mean and intensity ρ^{-1}. Optimally smoothed estimates can be obtained by running a Kalman filter in each of the causal and anticausal directions [34]. As discussed earlier, in general, we wish to compute only the most recent estimate (termed the "filtered estimate") $\hat{f}(s_1, s_2, \tau)$ from (4) for each $\tau \geq 0$. Such an estimate can be obtained by a single causal Kalman filter. To compute the filtered estimates, we discretize the first order Gauss–Markov dynamic equation as

$$\boldsymbol{f}(t) = \boldsymbol{f}(t-1) + \boldsymbol{q}(t), \qquad \boldsymbol{q}(t) \sim (0, \rho^{-1}I) \tag{9}$$

where the process noise $\boldsymbol{q}(t)$ is uncorrelated over time. This discrete dynamic model indicates that the optical flow evolves in time as the accumulation of a random perturbation at each time frame. Thus, the multiframe optical flow is formulated as a state estimation problem for the dynamic system whose dynamic equation is (9) and whose observation equation is given by the single-frame equation (6). State estimation for the dynamic system specified by (9) and (6) may be performed using the following implementation of the *information form* [32], [34] of the Kalman filter:

- *prediction stage*

$$\overline{\boldsymbol{L}}(t) = \rho I - \rho^2(\hat{\boldsymbol{L}}(t-1) + \rho I)^{-1} \tag{10}$$

$$\overline{\boldsymbol{f}}(t) = \hat{\boldsymbol{f}}(t-1) \tag{11}$$

$$\overline{\boldsymbol{z}}(t) = \overline{\boldsymbol{L}}(t)\overline{\boldsymbol{f}}(t) \tag{12}$$

- *update stage*

$$\hat{\boldsymbol{L}}(t) = \overline{\boldsymbol{L}}(t) + \boldsymbol{H}^T(t)\boldsymbol{W}(t)\boldsymbol{H}(t) + \mu_1\boldsymbol{S}_1^T\boldsymbol{S}_1 + \mu_2\boldsymbol{S}_2^T\boldsymbol{S}_2 \tag{13}$$

$$\hat{\boldsymbol{z}}(t) = \overline{\boldsymbol{z}}(t) + \boldsymbol{H}^T(t)\boldsymbol{W}(t)\boldsymbol{g}(t) \tag{14}$$

$$\hat{\boldsymbol{L}}(t)\hat{\boldsymbol{f}}(t) = \hat{\boldsymbol{z}}(t) \tag{15}$$

where $\overline{\boldsymbol{f}}(t)$ is the one-step predicted estimate, and $\hat{\boldsymbol{f}}(t)$ is the updated estimate using the new data available at time t. In addition $\overline{\boldsymbol{L}}(t)$ and $\hat{\boldsymbol{L}}(t)$ denote the predicted and updated information matrices, respectively.

The filtering algorithm (10)–(15) directly calculates the model parameters for the estimation error processes for the prediction and update steps at each time interval. In particular, the information matrices $\overline{\boldsymbol{L}}(t)$ and $\hat{\boldsymbol{L}}(t)$ specify the spatial models, which are analogous to the MRF model (8), that *implicitly* characterize the statistics of their respective error processes as

$$\overline{\boldsymbol{L}}(t)\overline{\boldsymbol{e}}(t) = \overline{\zeta}(t), \qquad \overline{\zeta}(t) \sim (0, \overline{\boldsymbol{L}}(t)) \tag{16}$$

$$\hat{\boldsymbol{L}}(t)\hat{\boldsymbol{e}}(t) = \hat{\zeta}(t), \qquad \hat{\zeta}(t) \sim (0, \hat{\boldsymbol{L}}(t)) \tag{17}$$

where $\overline{\boldsymbol{e}}(t)$ and $\hat{\boldsymbol{e}}(t)$ are predicted and updated estimation error processes, respectively.

The domains of support, or neighborhoods, associated with these spatial models deserve some attention. Recall that in the single-frame problem, the sparse "nearest-neighbor" structure of the information matrix $\boldsymbol{L}(t)$ is reflected in the compact neighborhood for the corresponding MRF model (8) and that this small and local support of the model facilitates the efficient solution of (7) for the estimate through iterative inversion. To understand how we may use these insights from the single-frame problem in the multiframe formulation, consider first the update stage (13)–(15). If $\overline{\boldsymbol{L}}(t)$ possesses a sparse and banded nearest-neighbor structure, then (13) preserves this structure in $\hat{\boldsymbol{L}}(t)$. In this case, (15) can be solved efficiently for the updated estimate $\hat{\boldsymbol{f}}(t)$ as this step would have *exactly* the same computational complexity as in the single-frame inversion step (7). Thus, preserving a nearest-neighbor structure in $\overline{\boldsymbol{L}}(t)$ is desirable from the computational standpoint. Unfortunately, the prediction step (10) will not preserve this structure, instead yielding a $\overline{\boldsymbol{L}}(t)$, which is a full matrix in general, even if $\hat{\boldsymbol{L}}(t-1)$ on the right-hand side is initially sparse. A full $\overline{\boldsymbol{L}}(t)$ will then lead to a full $\hat{\boldsymbol{L}}(t)$ by (13). Having such full information matrices makes the solution of (15) for the updated estimate computationally impractical. It also means that the corresponding spatial models would, in general, have a domain of support covering the entire image frame. Because of this lack of spatial locality arising in the general filtering equations (as reflected in the fullness of the information matrices), the

statistical properties of the multiframe estimates computed with the Kalman filter (10)–(15) do not have a compact MRF model representation. In Section V-A, however, we present a suboptimal Kalman filter that preserves a nearest-neighbor structure in the information matrices, leading to both a computationally efficient algorithm and an MRF-based statistical interpretation for the multiframe optical flow estimates.

V. Implementational Issues

A. Suboptimal Kalman filtering

A direct implementation of the optimal information Kalman filter (10)–(15) is impractical, as the number of pixels N in a frame of a typical image sequence is on the order of 10^4 to 10^6. The storage of the $O(N^2)$ elements of the information matrix as well as the inversion of the matrix in step (10) are particularly prohibitive to implement. Recall from the last section, however, that if $\overline{\boldsymbol{L}}(t)$ had a nearest-neighbor structure, this structure would be preserved in the information matrix through the rest of the stages of the filter. Thus, the key to our suboptimal filter is to force all the information matrices to have a nearest-neighbor structure through approximation of the prediction step (10).

From the viewpoint of the implicit statistical models (16) and (17), imposing a sparse structural constraint on the information matrix as above corresponds to constraining the support of the corresponding MRF models to be spatially local. In particular, if $\overline{\boldsymbol{L}}(t)$ and $\hat{\boldsymbol{L}}(t)$ are constrained to have a spatially local and symmetric nearest-neighbor structure, the estimation error models (16) and (17) now have representations as compact MRF models. Thus, our suboptimal filter propagates approximate, *reduced-order* models for the estimation error processes through the imposition of an MRF neighborhood of a fixed spatial extent on these processes.

As detailed in [26] and [41], such a reduced-order approximation may be obtained by expanding the matrix inverse on the right hand of (10) in a series as follows:

$$\overline{\boldsymbol{L}}(t) = \rho I - \rho^2(\Lambda^{-1} - \Lambda^{-1}\Omega\Lambda^{-1} + \Lambda^{-1}\Omega\Lambda^{-1}\Omega\Lambda^{-1} - \cdots) \tag{18}$$

where Λ is a block diagonal matrix whose 2×2 diagonal blocks are identical to the corresponding diagonal blocks of the matrix $\hat{\boldsymbol{L}}(t-1))+\rho I$, whereas $\Omega \equiv \hat{\boldsymbol{L}}(t-1)+\rho I-\Lambda$ is given by the remaining off-diagonal part of $\hat{\boldsymbol{L}}(t-1))+\rho I$. Note that Λ^{-1} is block diagonal. The series (18) may 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 the approximated matrix will become. Thus, there is a tradeoff between accuracy and computational efficiency. Our experience has shown that retaining only the first two terms as

$$\overline{\boldsymbol{L}}(t) = \rho I - \rho^2(\Lambda^{-1} - \Lambda^{-1}\Omega\Lambda^{-1}) \tag{19}$$

yields excellent results. Our near-optimal filter is obtained by replacing the optimal prediction step (10) by this two-term approximation. Unlike (10), the suboptimal prediction step (19) does indeed preserve the desired nearest neighbor structure in the (approximated) information matrix $\overline{\boldsymbol{L}}(t)$.

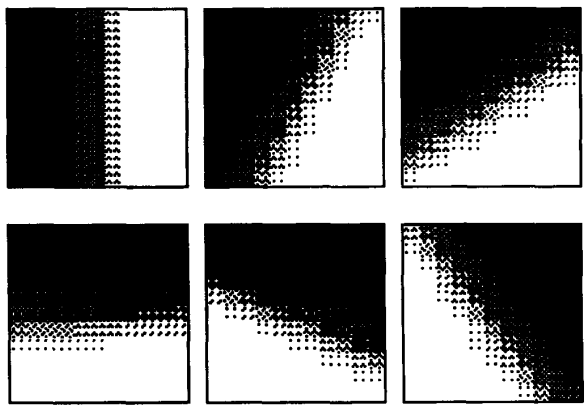


Fig. 1. Rotating ramp. Frames 0, 5, 10, 15, 20, and 25 are shown.

Propagating the information matrix in the suboptimal filter as in (13) and (19) 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 $O(N)$ nonzero 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.

B. Variance Computation

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

$$\hat{\boldsymbol{P}}^{(k+1)}(t) = \Lambda_L^{-1} - \Lambda_L^{-1}\Omega_L\hat{\boldsymbol{P}}^{(k)}(t) \tag{20}$$

where Λ_L is a block diagonal matrix whose 2×2 diagonal blocks are identical to the corresponding diagonal blocks of $\hat{\boldsymbol{L}}(t)$, and $\Omega_L \equiv \hat{\boldsymbol{L}}(t) - \Lambda_L$ is the remaining off-diagonal part of $\hat{\boldsymbol{L}}(t)$. This is a matrix version of the Jacobi iteration that is guaranteed to converge, i.e., $\hat{\boldsymbol{P}}^{(k)}(t) \rightarrow \hat{\boldsymbol{P}}(t)$ as $k \rightarrow \infty$ because $\hat{\boldsymbol{L}}(t)$ is positive definite [42]. We initialize $\hat{\boldsymbol{P}}^{(o)}(t) = \Lambda_L^{-1}$, which makes the recursion (20) equivalent (in the limit) to the series expansion used to invert a matrix in (18) [26].

The recursive scheme (20) is attractive in our suboptimal filtering scheme, where $\hat{\boldsymbol{L}}(t)$ is sparse at all times. Moreover, in practice, typically only certain elements of $\hat{\boldsymbol{P}}(t)$, namely, the diagonal elements representing the variances as well as the elements near the diagonal, are desired. The recursion (20) 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 $\hat{\boldsymbol{L}}(t)$, allows practical computation of variances.

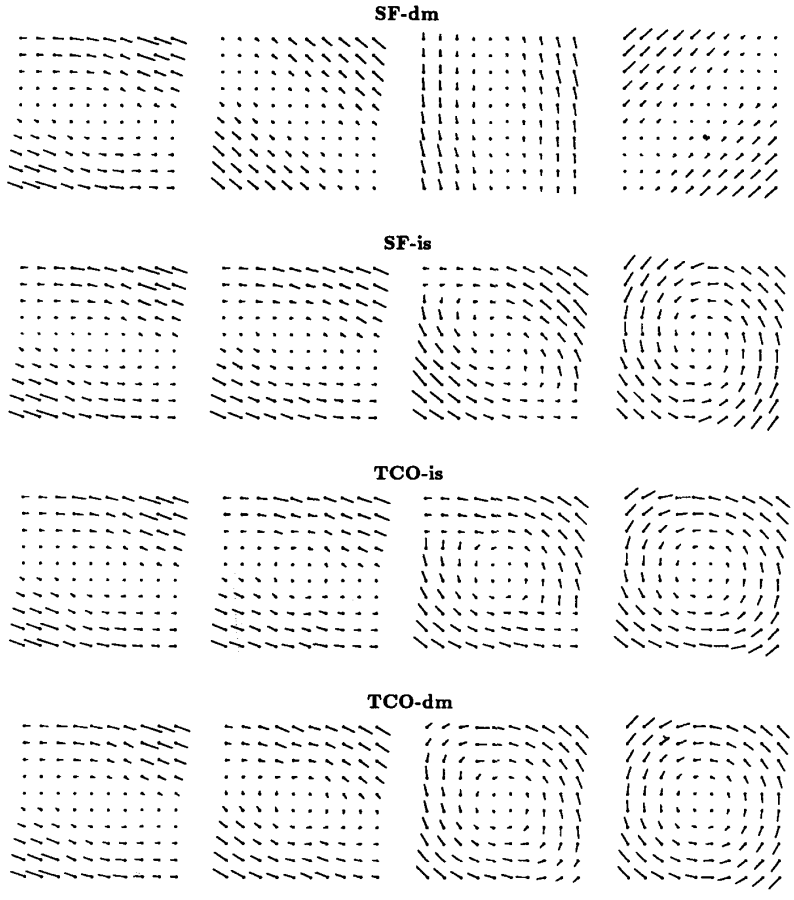


Fig. 2. Optical flow estimates for the rotating ramp example. The flow patterns at frames (from left to right) 0, 5, 15, and 25 are shown. The flow vectors are magnified by 1.5 for clarity.

C. Image Processing

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

$$E(\boldsymbol{s} + \Delta\boldsymbol{s}, t + \Delta t) - E(\boldsymbol{s}, t) = 0 \tag{21}$$

where $\boldsymbol{s} \equiv [s_1, s_2]^T$ and $\Delta\boldsymbol{s} \equiv (\Delta t)f$. By performing a Taylor series expansion of (21) we obtain

$$0 = \frac{\partial E}{\partial \boldsymbol{s}}(\Delta\boldsymbol{s}) + \frac{\partial E}{\partial t}(\Delta t) + \frac{\partial^2 E}{\partial \boldsymbol{s}\partial t}(\Delta\boldsymbol{s})(\Delta t) + \frac{1}{2}(\Delta\boldsymbol{s})^T \frac{\partial^2 E}{\partial \boldsymbol{s}^2}(\Delta\boldsymbol{s}) + \frac{(\Delta t)^2}{2}\frac{\partial^2 E}{\partial t^2} + \cdots \tag{22}$$

Then, by diving both sides by Δt, we see that (22) reduces to (2) only if all the second-order partial derivatives are zero

$$\frac{\partial^2 E}{\partial \boldsymbol{s}\partial t} = 0, \qquad \frac{\partial^2 E}{\partial \boldsymbol{s}^2} = 0, \qquad \text{and} \qquad \frac{\partial^2 E}{\partial t^2} = 0 \tag{23}$$

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 Section VI, presmoothing is implemented by averaging over 9×9 local subframes,[1] and improvements in accuracy (over the cases in which no presomoothing has been applied) of the optical flow estimates are observed. We further ensure the quality of the measurement by computing the second-order gradients $\partial^2 E/\partial \boldsymbol{s}\partial t, \partial^2 E/\partial \boldsymbol{s}^2$, and $\partial^2 E/\partial t^2$ 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 (see Section VI-C) that using the

[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.

weights

$$\nu(s_1, s_2, t) = \exp\left(-k\left\|\frac{\partial^2 E}{\partial \boldsymbol{s}\partial t}\right\|^2\right)$$
$$\equiv \exp\left(-k\left\|\left[\frac{\partial^2 E}{\partial \boldsymbol{s}_1\partial t}, \frac{\partial^2 E}{\partial \boldsymbol{s}_2\partial t}\right]\right\|^2\right) \quad (24)$$

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

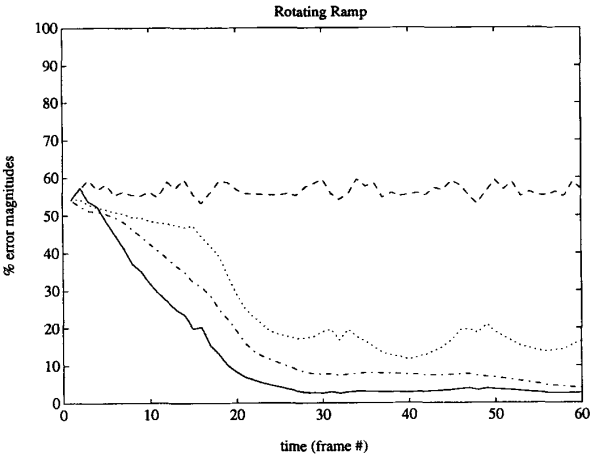
VI. Simulations

Synthetic image sequences of moving brightness patterns are processed by various multiframe 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:

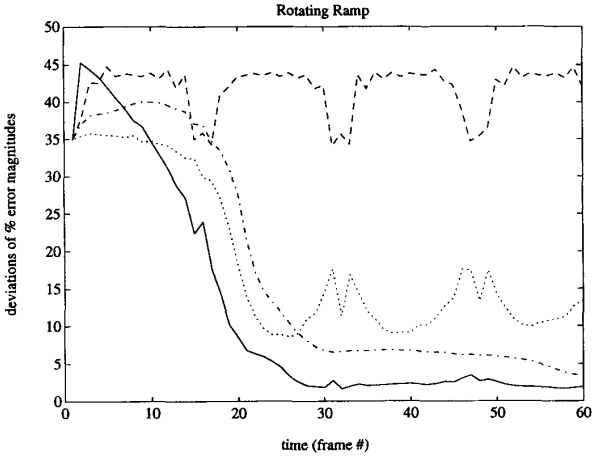
- **SF** (*Single Frame*): Each frame of optical flow is computed without any provision for temporal integration of data by solving the inversion problem (7) for $\hat{\boldsymbol{f}}(t)$. It corresponds to a discrete and single-frame version of the Horn and Schunck [1] formulation.
- **TCO** (*Temporal Coherence, Optimally Computed*): The flow estimates based on the temporal coherence constraint (9) are computed with the multiframe algorithm implemented as the *optimal* Kalman filter (10)–(15).
- **TCS** (*Temporal Coeherence, Suboptimally Computed*): This method is the approximate version of the **TCO** method: The prediction step (10) of the Kalman filter is approximated as (19).

Variants of these methods arise in different computational environments. Specifically, the inversion steps (7) (for **SF**) and (15) (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*): Direct matrix inversion provides us with the *exact* estimates. While its computational requirements are too large for typical optical flow problems, in one of the experiments to be presented, we have chosen to use a very small image frame so that we may perform direct inversion for comparison purposes.
- **ic** (*iterative inversion, iterations to convergence*): In practice, the inversion problems are solved iteratively. As in [1], we use Gauss–Seidel iterations in the experiments in this paper. This iterative solution should 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.



(a)

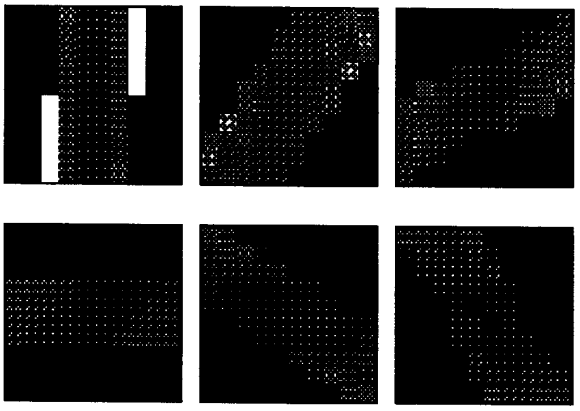


(b)

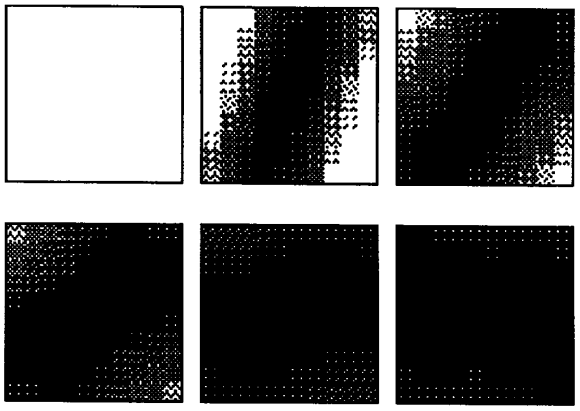
Fig. 3. (a) Average estimation errors and (b) their associated standard deviations in the rotating ramp example for the four methods: **SF-dm** (dashed line), **SF-is** (dotted line), **TCO-is** (dash-dot line), and **TCO-dm** (solid line).

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. In addition, 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.

Before proceeding, let us discuss the method **SF-is**. This method is the approach to multiframe optical flow estimation suggested by Horn and Schunck in [1]. At each time t, it uses the estimate from the *previous* frame $\hat{\boldsymbol{f}}(t-1)$ to initialize an iterative solution to (7) at the current time but then performs only *one* Gauss–Seidel iteration on this equation. Unlike the **SF-dm** or **SF-ic** method, therefore, this method *does* have some provision for propagating the estimates temporally



(a)



(b)

Fig. 4. 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.

(through the mechanism of the initial guess). 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.

A. Measurement Integration by Temporal Coherence

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, 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). Such difficulties can be circumvented by allowing for the use of information from adjacent image frames through the addition of a temporal coherence constraint, as demonstrated below.

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 rad. Fig. 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 five pixels wide. The pixel values are measured with floating-point accuracy without noise. The ramp is the only region in the image with nonzero 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: Fig. 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 $\boldsymbol{W}(t) = I$. 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 as well. 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: Fig. 3 displays the percentage average estimation error for each t

$$\frac{\|\hat{\boldsymbol{f}}(t) - \boldsymbol{f}(t)\|}{\|\boldsymbol{f}(t)\|} \times 100 \tag{25}$$

where $\boldsymbol{f}(t)$ is the true flow, and $\hat{\boldsymbol{f}}(t)$ 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 $\hat{\boldsymbol{f}}(t) - \boldsymbol{f}(t)$ 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, whereas 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

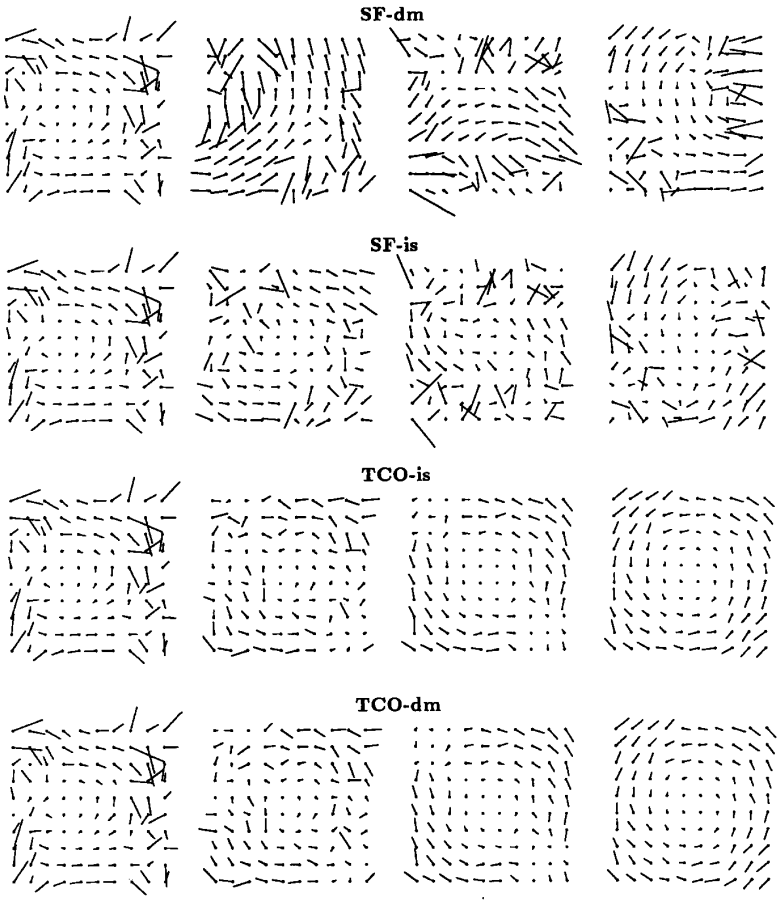


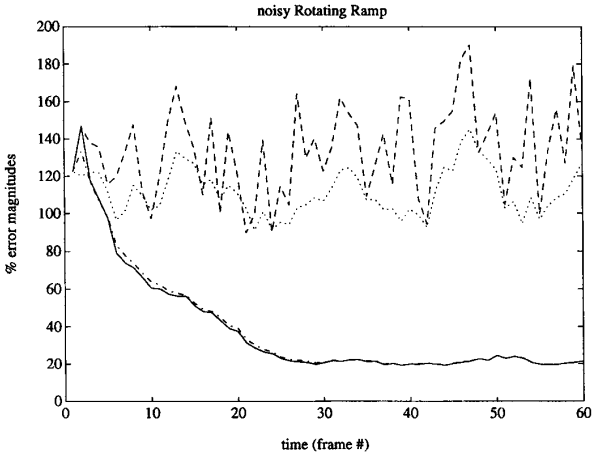
Fig. 5. Optical flow estimates for the *noisy* rotating ramp example. The flow patterns at frames (from left to right) 0, 5, 15, 25 are shown. The flow vectors are magnified by 1.5 for clarity.

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**.

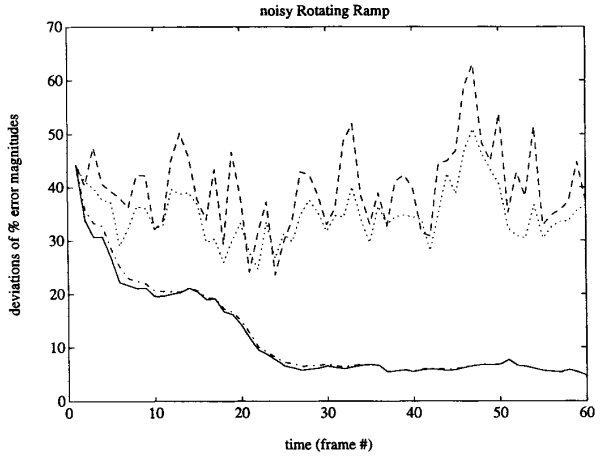
4) The Kalman gains and variances for the **TCO-dm** *method:* One can visualize the temporal integration process of the optimal Kalman filter in the **TCO-dm** method by observing the "images" of the magnitudes of the Kalman gains and variances. Fig. 4(a) shows the magnitudes of the Kalman gains at 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 Figs. 4(a) and 1 reveals that the Kalman gain is high where the image contrast is high. Fig. 4(b) 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.

B. 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. Fig. 5 displays the estimated flow fields, and Fig. 6 shows the estimation errors (25) 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 25th frame estimates in

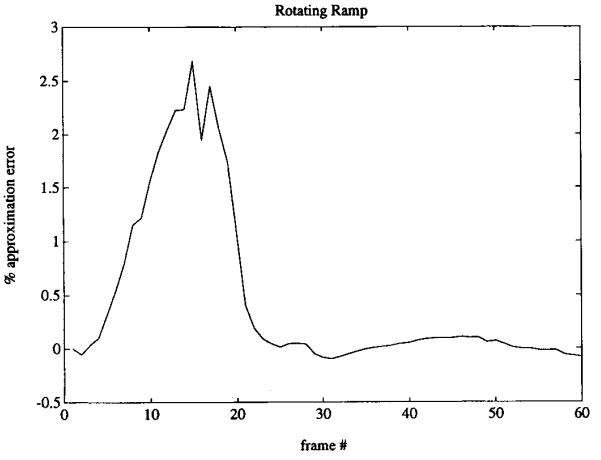


(a)

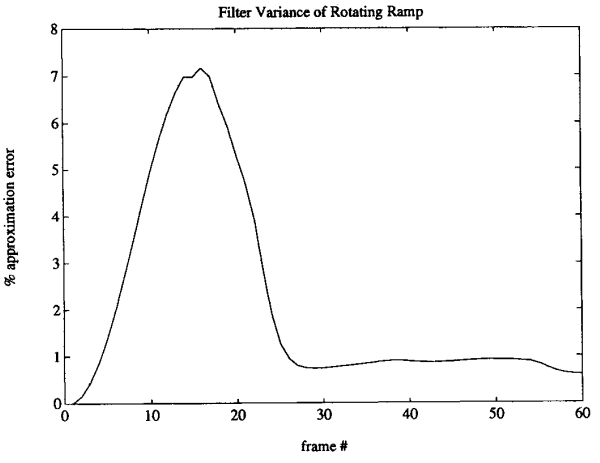


(b)

Fig. 6. (a) Estimation errors and (b) 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).



(a)



(b)

Fig. 7. Error introduced when the optimal Kalman filter in the **TCO-dm** method is approximated as in the **TCS-dm** method. The normalized approximation errors for the (a) estimates and (b) their standard deviations are shown.

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.

C. Temporal Coherence in the Approximate Filter

As discussed in Section V, 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 that demonstrate the efficacy of this near-optical **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 the 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 (25) 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 (see (19), (11)–(15)) 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

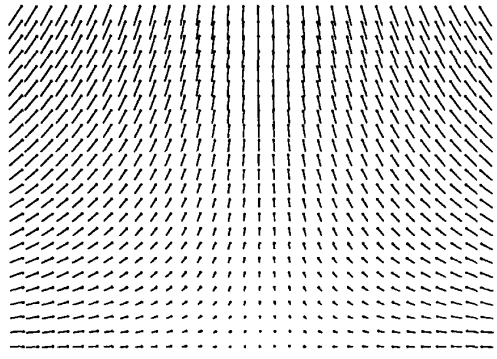


Fig. 8. True flow in the stagnation flow experiment. Every other flow vector along each axes is shown with a magnification factor of 4 for clarity. (Reprinted with permission of Kluwer Academic Publishers).

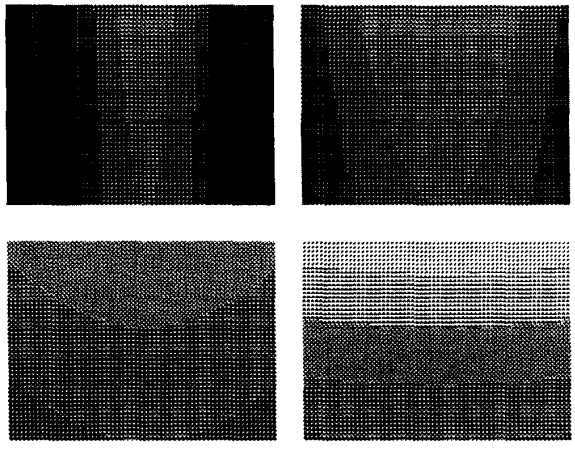


Fig. 9. Stagnation flow image sequence. Frames 0 and 7 (top row) as well as 14 and 21 (bottom row) are shown. (Reprinted with permission of Kluwer Academic Publishers).

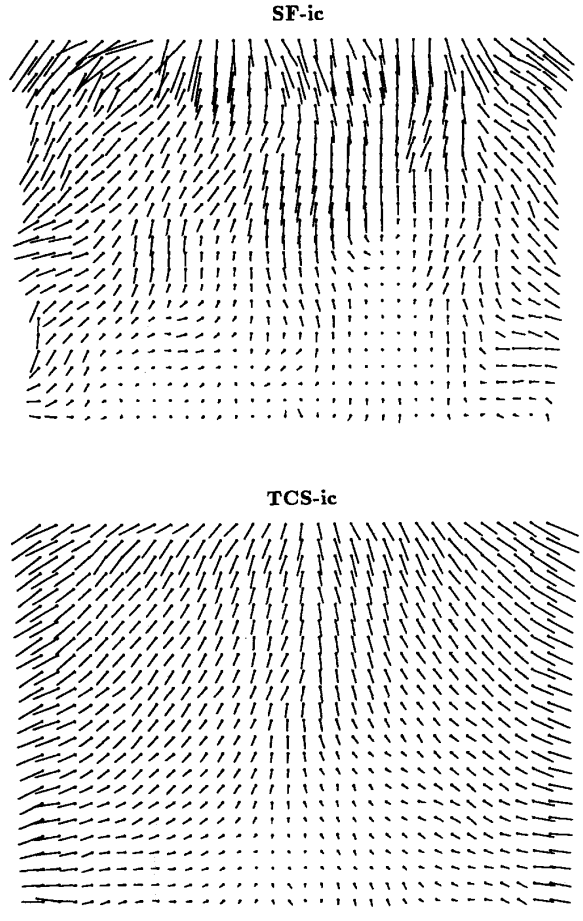


Fig. 10. Optical flow estimates for frame 18 of the stagnation flow sequence by the **SF-ic** and **TCS-ic** methods. (Reprinted with permission of Kluwer Academic Publishers.)

difference in percent normalized average estimation errors

$$\left(\frac{\|\hat{\boldsymbol{f}}_s(t)-\boldsymbol{f}(t)\|}{\|\boldsymbol{f}(t)\|}-\frac{\|\hat{\boldsymbol{f}}_o(t)-\boldsymbol{f}(t)\|}{\|\boldsymbol{f}(t)\|}\right)\times 100 \qquad (26)$$

where $\hat{\boldsymbol{f}}_s(t)$ and $\hat{\boldsymbol{f}}_o(t)$ are the estimates by the suboptimal and optimal filters, respectively, and $\boldsymbol{f}(t)$ is the true flow. Fig. 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

$$\frac{\|\sqrt{\boldsymbol{p}_s(t)}-\sqrt{\boldsymbol{p}_o(t)}\|}{\|\sqrt{\boldsymbol{p}_o(t)}\|}\times 100 \qquad (27)$$

comparing the variances (diagonals of the covariance matrices) $\boldsymbol{p}_s(t)$ and $\boldsymbol{p}_o(t)$ 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 thus far, which we will do next.

Experiment 2: Stagnation Flow: In this experiment,[2] the **SF-ic**, **TSC-ic**, **SF-is**, and **TSC-is** methods are used to estimate the motion of a nonrigid body.

1) The Image Sequence: This image sequence is based on a model of *stagnation flow* [46], i.e., the flow of fluid obstructed perpendicularly by a solid object. In particular, Fig. 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. A sequence of 64×48 images are synthesized based on such a velocity field.

[2] A portion of this experiment has appeared in [25] and is presented here with permission of Kluwer Academic Publishers.

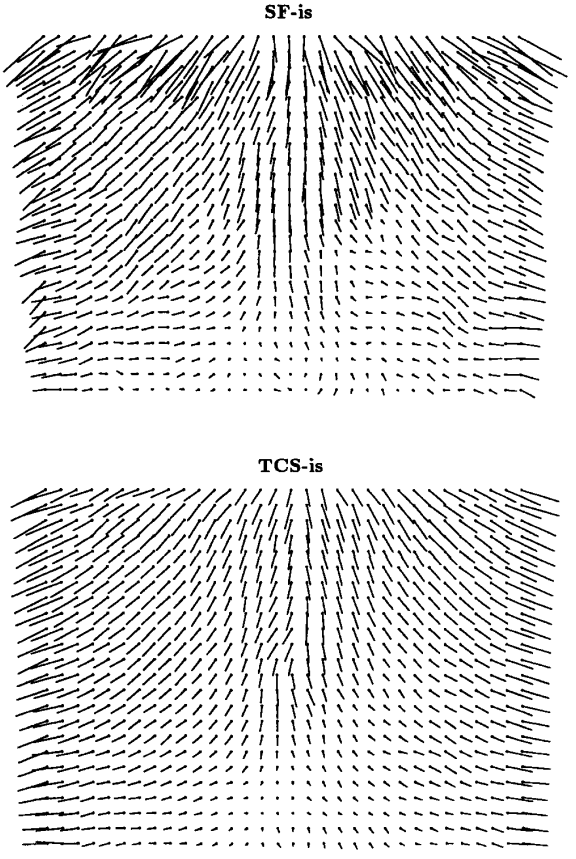


Fig. 11. Optical flow estimates for the frame 18 of the stagnation flow sequence by the **SF-is** method and **TCS-is** method. (Reprinted with permission of Kluwer Academic Publishers.)

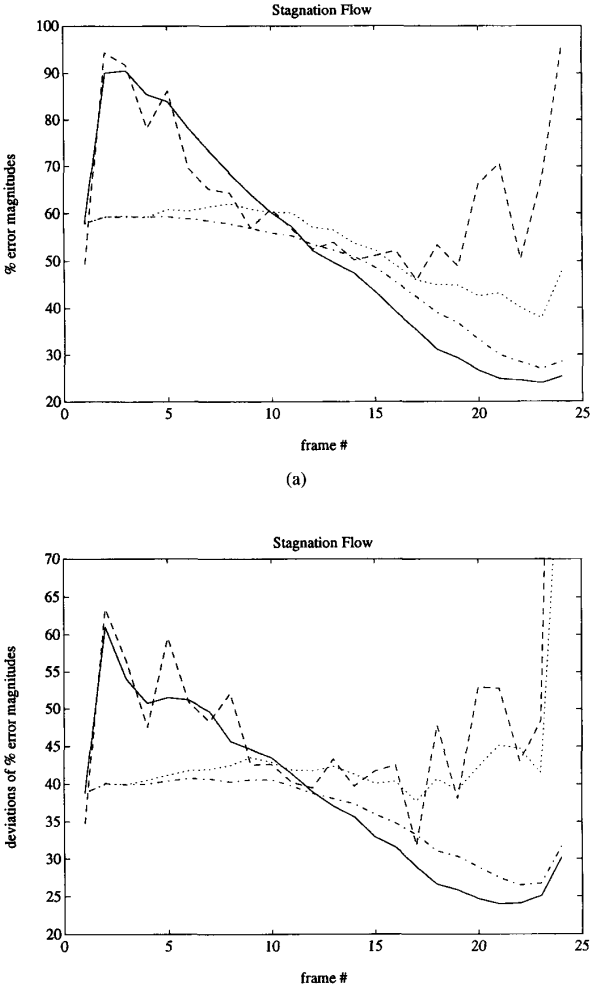


Fig. 12. (a) Estimation errors and (b) associated standard deviations by the **TCS-ic** (solid-line), **SF-ic** (dashed line), **SF-is** (dotted-line), and **TCS-is** (dash-dot line) methods for the stagnation flow experiment. (The estimation error plot reprinted with permission of Kluwer Academic Publishers.)

Fig. 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) Flow Estimates and Estimation Errors: As described in Section V-C, 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. Fig. 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, whereas the **TCS-ic** method has performed a reasonable reproduction of the flow in Fig. 8. The flows computed by the **SF-is** and **TCS-is** are shown in Fig. 11, which also displays the importance of temporal coherence in estimation. The average estimation errors and associated standard deviations for the four methods, which are shown in Fig. 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 Fig. 13. Note that both algorithms initialize each iterative session (except in the first frame) using the respective estimates from the previous frame. Fig. 13 indicates that the **TCS-ic** method requires progressively fewer iterations to compute the

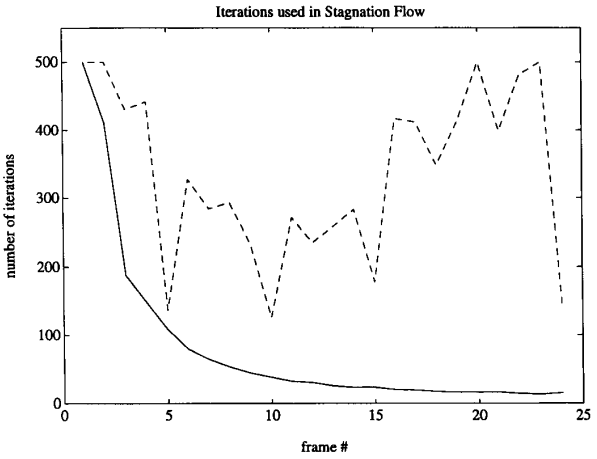


Fig. 13. Number of iterations used by the **TCS-ic** (solid-line) and **SF-ic** (dashed-line) methods for convergence of the estimates in the stagnation flow experiment.

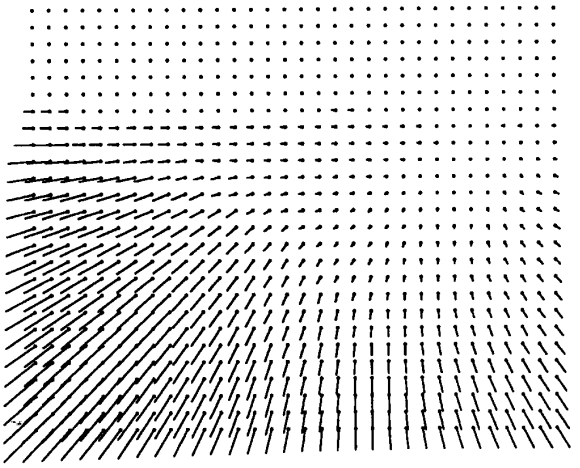


Fig. 15. True flow for the tenth frame in the Yosemite image sequence. The flow vectors are reduced by a factor of 2 for display purposes.

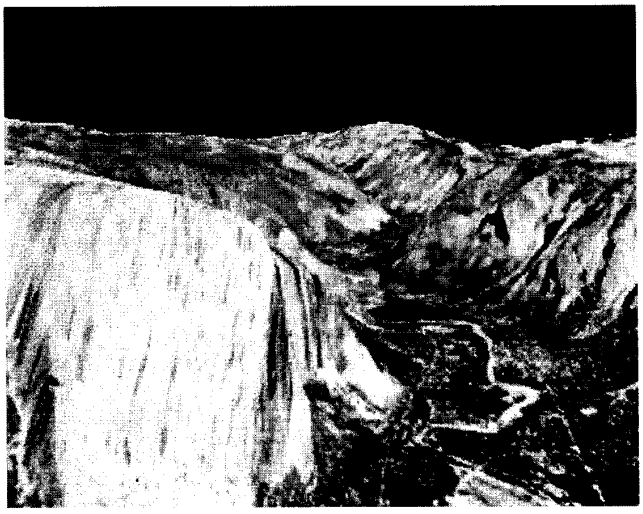


Fig. 14. Frame from the Yosemite image sequence.

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 (9) 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 (9), 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.[3] Fig. 14 shows a typical frame from the sequence. The frame size is 316×252, and the pixel values are quantized to 256 grey levels. Fig. 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 $\boldsymbol{W}(t)$, is computed using (24) with $k = \frac{1}{2}$. Compared with the uniform weighting $\boldsymbol{W}(t) = I$, 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 (24).

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 Fig. 16. The

[3]L. Quam of SRI International has produced the original image sequence.

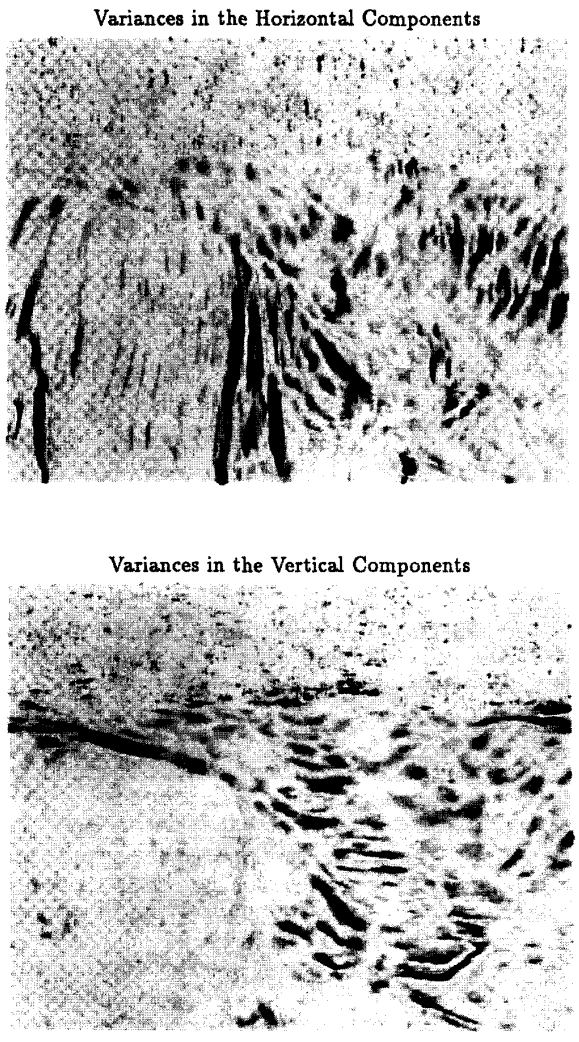


Fig. 16. Filter variances associated with the two components of each flow vector estimate for the second frame in the Yosemite image sequence.

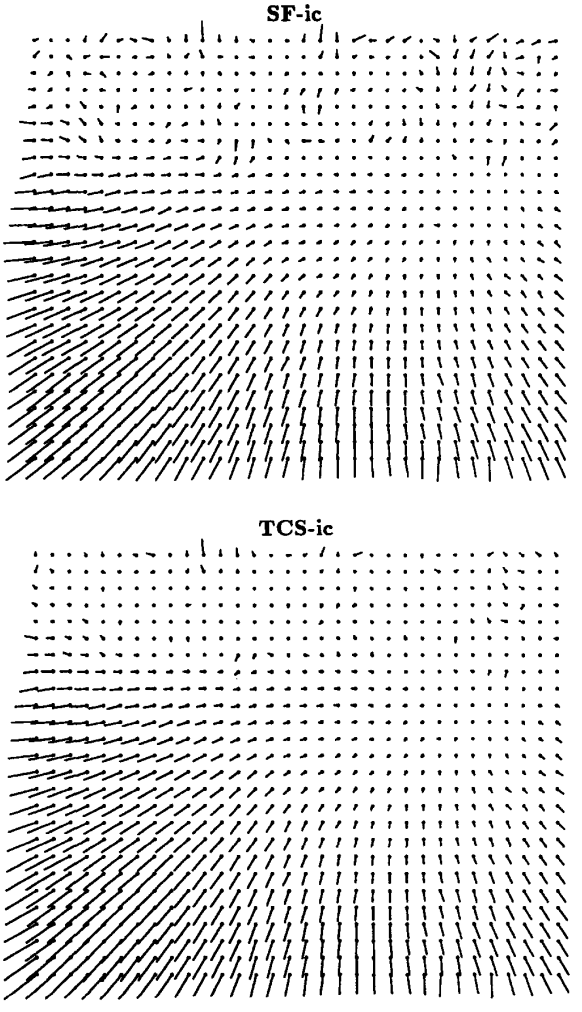
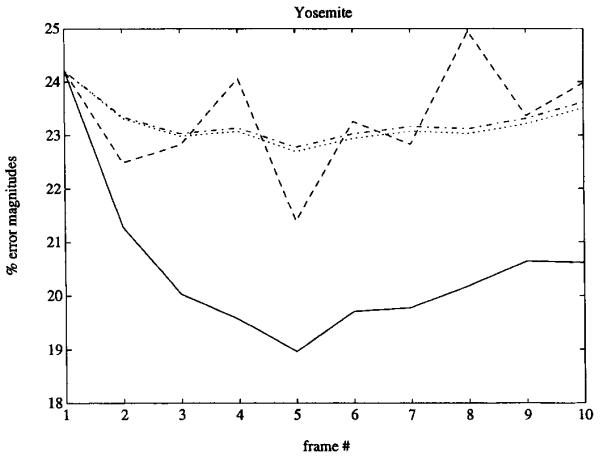


Fig. 17. Optical flow estimates for the tenth image frame by the **SF-ic** and **TCS-ic** methods. The flow vectors are reduced by a factor of 2 for display purposes.

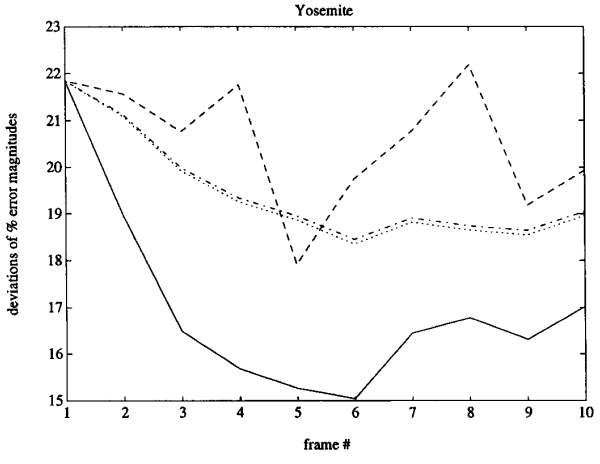
dark stripes in the displays are the pixel locations where the variances are low and where the image gradients are heavily relied on by the filter to estimate the flow field. As can be observed, the filter has taken advantage of the long and mostly linear gradients along the outlines and striations of 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 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. Fig. 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, Fig. 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 methods.

VII. Conclusion

We have demonstrated that the temporal coherence afforded by the use of the dynamic model (9) 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 multiframe 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



(a)



(b)

Fig. 18. (a) Estimation errors and (b) 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.

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 an time. This might involve studies on *piece-wise smooth* coherence constraints for the flow field (e.g., [35], [15]) and on spatio-temporal dynamic modeling of the discontinuity boundaries (e.g., [47]). In addition, 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 Intell.*, vol. 17, pp. 185–203, 1981.

[2] A. Meygret and M. Thonnat, "Segmentation of optical flow and 3D data for the interpretation of mobile objects," in *Third Int. Conf. Comput. Vision,* (Osaka, Japan), 1990, pp. 238–245.

[3] F. Heitz, P. Perez, E. Memin, and P. Bouthemy, "Parallel visual motion analysis using multiscale Markov random fields," in *Proc. Workshop Visual Motion.*(Princeton, NJ), 1991.

[4] H. C. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image," *Proc. Royal Soc. London B,* vol. 208, pp. 385–397, 1980.

[5] B. F. Buxton and H. Buxton, "Monocular depth perception from optical flow by space time signal processing," *Proc. Royal Soc. London B,* vol. 218, pp. 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 Trans. Patt. Anal. Machine Intell.,* vol. 8, pp. 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 *Proc. Third Int. Conf. Comput. Vision,* (Osaka, Japan), 1990, pp. 587–591.

[8] J. L. Prince and E. R. McVeigh, "Motion estimation from tagged MR image sequences," *IEEE Trans. Medical Imaging,* vol. 11, pp. 238–249, 1992.

[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," *J. Geophys. Res.,* vol. 91, no. 12, pp. 865–878, 1986.

[10] K. A. Kelly, "An inverse model for near-surface velocity from infrared images," *J. Phys. Oceanography,* vol. 19, pp. 1845–1864, 1989.

[11] A. N. Netravali and J. D. Robbins, "Motion-compensated television coding: Part I," *Bell Syst. Tech. J.,* vol. 58, pp. 631–670, 1979.

[12] ——, "Motion-compensated coding: Some new results," *Bell Syst. Tech. J.,* vol. 59, pp. 1735–1745, 1980.

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

[14] F. Glazer, "Hierachical motion detection," Ph.D. thesis, Univ. of Mass., 1987.

[15] H.-H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. PAMI-8, pp. 565–593, 1986.

[16] S. Uras, F. Girosi, A. Verri, and V. Torre, "A computational approach to motion perception," *Biolog. Cybern.,* vol. 60, pp. 79–97, 1988.

[17] P. Anandan, "Measuring visual motion from image sequences," Ph.D. thesis, Univ. of Mass, Amherst, 1987.

[18] D. J. Heeger, "Optical flow using spatiotemporal filters," *Int. J. Comput. Vision,* vol. 1, pp. 279–302, 1988.

[19] D. J. Fleet and A. D. Jepson, "Computation of component image velocity from local phase information," *Int. J. Comput. Vision,* vol. 5, pp. 77–104, 1990.

[20] 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, MI), 1988, pp. 717–723.

[21] J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt, "Performance of optical flow techniques," Rep. no. 299, Dept. of Comput. Sci., Univ. of Western Ontario, 1991.

[22] J. J. Little and A. Verri, "Analysis of differential and matching methods for optical flow," in *Proc. Workshop Visual Motion,* (Irvine, CA), 1989, pp. 173–180.

[23] E. C. Hildreth, "Computations underlying the measurement of visual motion," *Artif. Intell.,* vol. 23, pp. 309–354, 1984.

[24] R. Szeliski, *Baysian Modeling of Uncertainty in Low-level Vision.* Norwell, MA: Kluwer, 1989.

[25] T. M. Chin, M. R. Luettgen, W. C. Karl, and A. S. Willsky, "An estimation theoretic perspective on image processing and the calculation of optical flow," in *Motion Analysis and Image Sequence Processing,* I. Sezan and R. Lagendijk, Eds. Norwell, MA: Kluwer, 1993.

[26] T. M. Chin, W. C. Karl, and A. S. Willsky, "Sequential filtering for multi-frame visual reconstruction," *Signal Processing,* vol. 28, pp. 311–333, 1992.

[27] J. W. Woods and C. H. Radewan, "Kalman filtering in two dimensions," *IEEE Trans. Inform. Theory,* vol. IT-23, pp. 473–482, 1977; see also "Correction" in IT-25, pp. 628–629.

[28] M. S. Murphy and L. M. Silverman, "Image model representation and line-by-line recursive restoration," *IEEE Trans. Automat. Contr.,* vol. AC-23, pp. 809–816, 1978.

[29] J. W. Woods, "Two-dimensional Kalman filtering," in *Two-Dimensional, Digital Signal Processing I,* (T. S. Huang, Ed.) New York: Springer-Verlag, 1981, pp. 155–205.

[30] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proc. IEEE,* vol. 76, pp. 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 *Proc. Workshop Visual Motion,* (Irvine, CA), 1989, pp. 148–155.
[32] B. D. O. Anderson and J. B. Moore, *Optimal Filtering.* Englewood Cliffs, NJ: Prentice-Hall, 1979.
[33] A. Gelb, Ed., *Applied Optimal Estimation.* Cambridge, MA: MIT Press, 1974.
[34] F. L. Lewis, *Optimal Estimation.* New York: Wiley, 1986.
[35] M. J. Black and P. Anandan, "A model for the detection of motion over time," in *Third Int. Conf. Comput. Vision,* (Osaka, Japan), 1990, pp. 33–37.
[36] A. Singh, "Incremental estimation of image-flow using a Kalman filter," in *Proc. Workshop Visual Motion* (Princeton, NJ), 1991, pp. 36–43.
[37] B. C. Levy, M. B. Adams, and A. S. Willsky, "Solution and linear estimation of 2-D nearest-neighbor models," *Proc. IEEE,* vol. 78, pp. 627–641, 1990.
[38] D. Terzopoulos, "Image analysis using multigrid relaxation models," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. PAMI-8, pp. 129–139, 1986.
[39] B. C. Levy, R. Frezza, and A. J. Krener, "Modeling and estimation of discrete-time Gaussian reciprocal processes," *IEEE Trans. Automat. Contr.,* vol. AC-35, pp. 1013–1023, 1990.
[40] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. PAMI-6, pp. 721–741, 1984.
[41] T. M. Chin, "Dynamic estimation in computational vision," Ph.D. thesis, Mass. Inst. Technol., Cambridge, MA, 1991.
[42] G. H. Golub and C. F. van Loan, *Matrix Computations.* Baltimore, MD: The Johns Hopkins University Press, 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 Trans. Patt. Anal. Machine Intell.,* vol. PAMI-9, pp. 229–244, 1987.
[44] D. M. Martinez, "Model-based motion estimation and its application to restoration and interpolation of motion pictures," Ph.D. thesis, Mass. Inst. Technol., 1986.
[45] M. K. Brown and B. Shahraray, "Robust depth estimation from optical flow," in *Proc. Comput. Vision Patt. Recogn.* (Tampa, FL), 1988, pp. 641–650.
[46] M. C. Potter and J. F. Foss, *Fluid Mechanics.* Okemos, MI: Great Lakes, 1982.
[47] 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 Phys. Oceanography),* 1993.

Toshio M. Chin (S'87) was born near Kobe, Japan, in 1960. He received the B.S. and M.S. degrees in electrical engineering from the University of California at Berkeley, in 1982 and 1983, respectively, was with the Computer Vision and Image Processing Laboratory, University of California at Davis from 1984–1985, was awarded a Harvard/M.I.T. Health Science and Technology Fellowship in 1986, and received the E.E. degree in 1990 and the Ph.D. degree in 1992 in electrical engineering from the Massachusetts Institute of Technology. He is currently a postdoctoral associate with the Meteorology and Physical Oceanography Department, University of Miami.

His research interests are in multidimensional signal processing, with emphasis in efficient numerical algorithms and estimation problems involving motion.

William C. Karl received the Ph.D. degree in electrical engineering and computer science in 1991 from the Massachusetts Institute of Technology, Cambridge, where he also received the S.M., E.E., and S.B. degrees.

He has held the position of Staff Research Scientist with the Brown-Harvard-M.I.T. Center for Intelligent Control Systems and the M.I.T. Laboratory for Information and Decision Systems since 1992. In 1993, he was organizer and chair of the "Geometry and Estimation" session of the Conference on Information Sciences and Systems at Johns Hopkins University. His research interests are in the areas of multidimensional and multiscale signal and image processing and estimation, particularly applied to geometrically oriented problems.

Alan S. Willsky (S'70–M'73–SM'82–F'86) received both the S.B. degree and the Ph.D. degree from the Massachusetts Institute of Technology (M.I.T.) in 1969 and 1973, respectively.

He joined the M.I.T. faculty in 1973, and his present position is Professor of Electrical Engineering. From 1974 to 1981, he served as Assistant Director of the M.I.T. Laboratory for Information and Decision Systems. He is also a founder and member of the board of directors of Alphatech, Inc.

His present research interests are in problems involving multidimensional and multiresolution estimation and imaging, discrete-event systems, and the asymptotic analysis of control and estimation systems.

In 1975, Dr. Willsky received the Donald P. Eckman Award from the American Automatic Control Council. He has held visiting positions at Imperial College, London, L'Université de Paris-Sud, and the Institut de Recherche en Informatique et Systèmes Aléatoires in Rennes, France. He was program chairman for the 17th IEEE Conference on Decision and Control, has been an associate editor of several journals including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, has served as a member of the Board of Governors and Vice President for Technical Affairs of the IEEE Control Systems Society, was program chairman for the *1981 Bilateral Seminar on Control Systems* held in the People's Republic of China, and was special guest editor of the 1992 special issue of the IEEE TRANSACTIONS ON INFORMATION THEORY on wavelet transforms and multiresolution signal analysis. In addition, in 1988, he was made a Distinguished Member of the IEEE Control Systems Society. In addition, Dr. Willsky has given several plenary lectures at major scientific meetings including the 20th IEEE Conference on Decision and Control, the 1991 IEEE International Conference on Systems Engineering, the SIAM Conference on Applied Linear Algebra in 1991, and the 1992 Inaugural Workshop for the National Centre for Robust and Adaptive Systems, Canberra, Australia. He is the author of the research monograph *Digital Signal Processing and Control and Estimation Theory* and is co-author of the undergraduate text *Signals and Systems.* He was awarded the 1979 Alfred Noble Prize by the ASCE and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE for a paper excerpted from his monograph.