July 1992 LIDS-P-2117

An Estimation Theoretic Perspective on Image Processing and the Calculation of Optical Flow

T.M. Chin
RSMAS, U. of Miami, Miami FL, USA

M.R. Luettgen
M.I.T., Cambridge MA, USA

W.C. Karl
M.I.T., Cambridge MA, USA

A.S. Willsky
M.I.T., Cambridge MA, USA

2.1 Introduction

Many problems of image processing and image sequence analysis involve both great computational complexity and the accommodation of noise and uncertainty through the indirect observation of quantities of interest. In this chapter we describe several aspects of an estimation theoretic approach to such problems. The vehicle for our development is the estimation of the apparent velocity field of a sequence of images. This apparent velocity field, known as the optical flow, appears as an important quantity in both the qualitative and quantitative analysis of image sequences. For example, knowledge of the optical flow is used in the detection of object boundaries and the segmentation of visual scenes [1, 2], the derivation of 3-D motion and structure [3, 4], and the compression of image sequences for efficient transmission [5, 6].

We take an estimation-theoretic perspective to the computation of optical

flow, using and extending the formulation of Rougee et al. [7, 8] in both the temporal and spatial directions. In particular, we use model-based interpretations of the various components arising in the estimation theoretic setting to allow us to develop novel extensions to existing approaches. First, we consider the imposition of a *temporal coherence* to the flow obtained by modeling the evolution of the vector optical flow process with a linear state equation and then applying a recursive Kalman filter to the observations obtained from the image sequence. The classical (Horn and Schunk [9]) formulation of the optical flow estimation problem contains no such formal requirement of temporal coherence. The inclusion of such a constraint allows the reliable and robust estimation of optical flow under conditions difficult for the classical approach. For example, in situations where a single image pair contains insufficient information to recover the flow field due to the"aperture problem," the integration of observations over a longer time frame can yield reasonable results.

Applications of Kalman filtering to various formulations of optical flow estimation [10, 11] as well as to other low-level reconstruction problems in computational vision [12] have been proposed. In these previous approaches, however, the apparently computationally daunting task of implementing the Kalman filtering equations, and in particular the error covariance equations, on even moderately-sized images resulted in the use of drastically simplified and suboptimal filter specifications. 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 [13, 14, 15] 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 the Kalman filter than those previously reported. This approximation is based on the propagation of approximate local *models* of the estimation error covariance. These results provide, to our knowledge, the first implementation of the *complete* Kalman filtering equations for space-time problems of this scale, and the only example of successful, near optimal, propagation of covariance matrices of this size. The mathematical details of our approximation techniques can be found in [16] in the more general context of low-level visual reconstruction.

Second, we use the observation that both the single and multi-frame problems can be formulated as spatial estimation problems, wherein sets of observations are fused with prior spatial field models, to motivate the use of a recently developed class of *multiscale* statistical models in their solution. What makes these multiscale field models especially interesting is 1) that there exist extremely efficient, multigrid-type estimation algorithms based on them and 2) that a large number of degrees of freedom exist in their specification, allowing them to approximate a wide range of different flows, including, as least conceptually, any Markov Random Field based flow. Together, these qualities imply that the utilization of such multiscale spatial models for spatial estimation problems, and in particular for the optical flow problem, provides a flexible yet *extremely* ef-

ficient estimation framework. Preliminary examples of our results are provided showing factors of 10-100 computational improvment over conventional methods. Finally, such models provide multiscale representations of the flow field and, though we have not used it here, also provide the possiblility of optimal integration of *multiscale measurements*.

In this chapter we focus on a particular image processing problem, namely the computation of optical flow. However, the model-based approaches used here are more generally applicable to the wide range of space-time estimation problems arising in image sequence processing.

2.2 Optical Flow Estimation

2.2.1 Single-Frame Formulation

The 2-D vector field of the apparent motion of brightness patterns in an image is referred to as the optical flow [9]. One commonly used way to obtain information about the optical flow field $x(z_1, z_2, t) \equiv [dz_1/dt, dz_2/dt]^T$ at a given point in space (z_1, z_2) and time t was presented by Horn and Schunck in [9]. This approach is based on the assumption that changes in scene brightness in the image sequence are due only to motion. This assumption leads to the so called *brightness constraint equation* [9]:

$$0 = \frac{d}{dt}E(z_1, z_2, t) = \frac{\partial}{\partial t}E(z_1, z_2, t) + \nabla E(z_1, z_2, t) \cdot x(z_1, z_2, t) \qquad (2.1)$$

where $E(z_1, z_2, t)$ is the image intensity as a function of time and space and $\nabla E = [\partial E/\partial z_1, \partial E/\partial z_2]$, is the gradient of the image intensity.

The brightness constraint equation (2.1) does not completely specify the flow field since it provides only one linear constraint for the two unknown components of $x(z_1, z_2, t)$ at each point. This is usually referred to as the aperture problem [9]. One way to obtain a unique solution is to *regularize* the problem by imposing an additional *smoothness constraint*. Specifically, one formulates the following optimization problem [9]:

$$\underset{x(z_1,z_2,t)}{\operatorname{argmin}} \iint \nu \, \| \frac{\partial}{\partial t}E(z_1, z_2, t) + \nabla E(z_1, z_2, t) \cdot x \|^2 + \|\nabla x\|^2 \, dz_1 dz_2 \qquad (2.2)$$

The smoothness constraint is captured by the second term which penalizes large gradients in the optical flow and is necessary to make the formulation mathematically well-posed [17]. This term also represents 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 [18]. The constant $\nu(z_1, z_2, t)$ allows one to tradeoff between the relative importance in the cost function of the brightness and smoothness constraint terms.

Before proceeding let us analyze the smoothness constraint in more detail. Note that the penalty associated with the smoothness constraint term in (2.2) is equal to the integral of the squared norm of the field gradient over the image

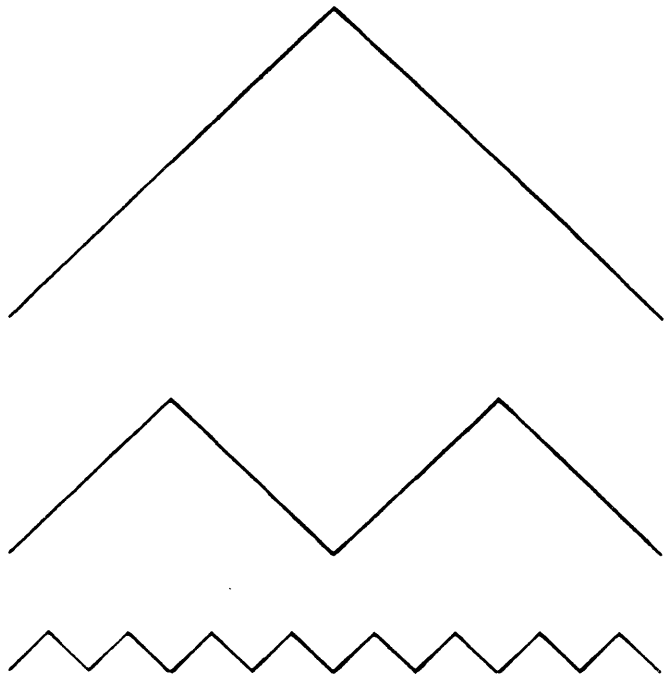

Figure 2.1: Depiction of three fields which are equally favored by the smoothness constraint, illustrating how this penalty provides a fractal prior model for the optical flow.

plane. In a one-dimensional context, such a constraint would penalize each of the (one-dimensional) fields in Figure 2.1 equally. Intuitively, the smoothness constraint has a fractal nature, and in fact this can be demonstrated in a much more precise sense, as we show in Section 2.3.1.

2.2.2 Multi-Frame Formulation

The formulation (2.2) processes the data (i.e. the gradients of the image intensity) a frame at a time, yielding flow estimates independently over time. The imposition of *temporal coherence* [19] to the flow field can be considered in addition to the spatial coherence enforced by (2.2) 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 of (2.2) [20, 12]. In particular, we find the flow field $\hat{x}(z_1, z_2, \tau)$ which is the solution to the following problem at time $t = \tau$:

$$\underset{x(z_1,z_2,t)}{\operatorname{argmin}} \int_0^\tau \iint_D \nu \left\| \frac{\partial E}{\partial t} + \nabla E(z_1, z_2, t) \cdot x \right\|^2 + \|\nabla x\|^2 + \rho \left\| \frac{\partial}{\partial t} x \right\|^2 ds\, dt. \quad (2.3)$$

The multi-frame formulation (2.3) is obtained from the single-frame formulation (2.2) by the addition of a quadratic term involving the first order temporal

derivative. Note that the full solution to the optimization problem (2.3) leads to a reconstructed space-time field $\hat{x}(z_1, z_2, t)$ in which the reconstruction at any time takes advantage of all available constraints over the entire time interval $0 \leq t \leq \tau$. We are only interested in the value of the solution to (2.3) at the current time $t = \tau$, corresponding to the best causal or filtered estimate of the flow field. 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 (2.2), making direct solution of the optimization problem (2.3) prohibitive.

2.3 Discretization and Probabilistic Interpretation

In the first part of this section we present a discrete formulation of the single-frame optical flow problem. In this context we develop two *model-based* interpretations of the single-frame problem which will be central to the results of Sections 2.4 and 2.5. In particular, we illustrate 1) how the smoothness constraint can be interpreted as a prior probabilistic spatial model for the flow field and 2) how the inverse of the covariance of the field estimate may be naturally interpreted as another spatial model, this time for the posterior estimation error of the field. The second part of the section shows how we may model the temporal coherence constraint of (2.3) by a discrete dynamic equation, which may then be coupled with a set of observation equations obtained from the single-frame case to yield an equivalent state estimation problem.

2.3.1 Single-Frame Case

In practice, brightness measurements are only available over a discrete set of points in space and time. Thus, the temporal and spatial derivative terms must be approximated with finite differences, and the optical flow is only estimated on a discrete space-time grid. There are a number of important issues which arise due to the discretization which we do not discuss here: we refer the reader to [21] for a detailed treatment. We will assume here that we have normalized the space-time coordinates so that the optical flow is to be estimated on the set of integers $(z_1, z_2) \in \{(i,j) | i,j \in \{1, \cdots, 2^M\}\}$, where M is also an integer, so the total number of image points is $N \equiv 4^M$. The assumption that the grid is square and that the number of rows is equal to a power of two makes the development of Section 2.5 easier, but is not essential. Now at grid point (i,j) at time t let us define $y(i,j) \equiv (-\partial E(i,j,t)/\partial t)$ to be the measured temporal brightness derivative, $x(i,j)$ to be the desired optical flow vector, and $C(i,j) = \nabla E(i,j,t)$ to be the spatial gradient of the image brightness (also measured). Then we may write the brightness constraint (2.1) at the point (i,j) at time t as:

$$y(i,j) = C(i,j)\,x(i,j) \tag{2.4}$$

The brightness constraints (2.4) at all grid points can now be grouped into one large set of linear equations to concisely capture the optical flow information contained in the image sequence. Let $\mathbf{x}(t)$ be a vector containing the optical flow values $x(i,j)$ at all the grid points at time t (using, say, a lexicographic ordering) and $\mathbf{y}(t)$ be the associated vector of the samples $y(i,j)$. Similarly, let $\mathbf{C}(t)$ and $\mathbf{W}(t)$ be block diagonal matrices whose diagonal elements are the samples $c(i,j)$ and $w(i,j)$, respectively, taken in the same order at time t. A discrete version of the single-frame formulation (2.2) is then given by:

$$\hat{\mathbf{x}}_{SC}(t) = \underset{\mathbf{x}(t)}{\operatorname{argmin}} \left\{ \|\mathbf{y}(t) - \mathbf{C}(t)\mathbf{x}(t)\|^2_{\mathbf{W}(t)} + \|\mathbf{S}\mathbf{x}(t)\|^2_I \right\} \tag{2.5}$$

where $\|\mathbf{x}\|^2_{\mathbf{W}}$ denotes the weighted norm $\mathbf{x}^T\mathbf{W}\mathbf{x}$. I represents the identity matrix, and $\mathbf{S}$ is the matrix first-order spatial difference operator. Note that the discrete nature of the problem alluded to above implies that we must actually approximate the samples $-\partial E(i,j,t)/\partial t$ and $\nabla E(i,j,t)$, and thus $\mathbf{C}$ and $\mathbf{y}$ by finite differences. The spatially varying entries of $\mathbf{W}$ can actually be used to reflect our confidence in these approximations. We refer the reader to [21] for further details.

Spatial Models

An estimation-theoretic formulation of the optimization problem in (2.5) can now be developed, and we will use it to show that the *statistically optimal* estimate of the optical flow, given a particular set of measurements, is identical to the smoothness constraint solution given in (2.5). Specifically, solving the quadratic minimization problem (2.5) is equivalent to solving a maximum likelihood (ML) estimation problem [15] for $\mathbf{x}(t)$ with the following *observation equations*:

$$\mathbf{y}(t) = \mathbf{C}(t)\,\mathbf{x}(t) + \mathbf{r}_1(t) \qquad \mathbf{r}_1(t) \sim \left(\mathbf{0}, \mathbf{W}^{-1}(t) \right) \tag{2.6}$$

$$0 = \mathbf{S}\,\mathbf{x}(t) + \mathbf{r}_2(t) \qquad \mathbf{r}_2(t) \sim (\mathbf{0}, I) \tag{2.7}$$

where we have used the notation $\mathbf{x} \sim (\mathbf{m}, \mathbf{H})$ to denote a Gaussian random vector $\mathbf{x}$ whose mean and covariance are $\mathbf{m}$ and $\mathbf{H}$, respectively, so $\mathbf{r}(t) \equiv [\mathbf{r}_1(t), \mathbf{r}_2(t)]^T$ is a zero-mean Gaussian random noise process. Thus, the maximum likelihood problem formulation results in the same solution as the smoothness constraint formulation when $\mathbf{S}$ is used to define an additional set of noisy measurements.

By formulating the problem in this estimation-theoretic framework, we can use (2.7) to interpret the smoothness constraint as a prior probabilistic spatial *model* for the flow field. Specifically, we can rewrite (2.7) as:

$$\mathbf{S}\mathbf{x}(t) = -\mathbf{r}_2(t) \tag{2.8}$$

Recalling that $\mathbf{S}$ is an approximation to the gradient operator, we see that (2.8) is nothing more than a spatial difference equation model for $\mathbf{x}(t)$ driven by the spatial white noise field $\mathbf{r}_2(t)$. In particular, this prior model represents the optical flow field as composed of independent, two-dimensional Brownian

motions[1] [7, 8]. Then, the statistically optimal estimate of the flow field, given the measurements (2.6) and the Brownian motion prior model, is the same as the optical flow estimate given by (2.5). The estimation-theoretic interpretation simply allows us to interpret the smoothness constraint as a Brownian motion model. In one-dimension, Brownian motion is a statistically self-similar, fractal process with a $1/f^2$ generalized spectrum [22], and for this reason the smoothness constraint is often referred to as a "fractal prior" [12]. We will return to this interpretation in Section 2.5 where we discuss a multiscale modeling approach to the single frame problem. In particular, we will replace the prior model (2.8) by a similar but multiscale prior model, which leads to dramatic computational savings.

Next, let us consider another model based interpretation of the single-frame problem (2.5) that will be useful in treating the multi-frame problem. The ML estimate for the optical flow, $\hat{\mathbf{x}}(t)$, based on the measurements (2.6),(2.7) is obtained as the solution of the following inverse problem:

$$\left(\mathbf{C}^T(t)\mathbf{W}(t)\mathbf{C}(t) + \mathbf{S}^T\mathbf{S}\right)\hat{\mathbf{x}}(t) = \mathbf{C}^T(t)\mathbf{W}(t)\mathbf{y}(t) \tag{2.9}$$

The equations in (2.9) represent a discrete version of the coupled Poisson equations of the Horn and Schunck formulation. The matrix operator

$$\mathbf{L}(t) = \left(\mathbf{C}^T(t)\mathbf{W}(t)\mathbf{C}(t) + \mathbf{S}^T\mathbf{S}\right) \tag{2.10}$$

on the left hand side of (2.9) has a sparse, *nearest neighbor* (a nested block tri-diagonal) structure [23], whose sparseness enables us to use efficient iterative procedures, such as multigrid methods [24], in the solution of (2.9). Also, this sparse matrix corresponds to the *information matrix* (the inverse of the covariance matrix) associated with the posterior estimation error $\mathbf{d}(t) \equiv \mathbf{x}(t) - \hat{\mathbf{x}}(t)$. In particular $\mathbf{L}(t)$ can naturally be considered to specify an implicit Markov Random Field model for the estimation error process $\mathbf{d}(t)$ of the following form:

$$\mathbf{L}(t)\,\mathbf{d}(t) = \zeta(t), \qquad \zeta(t) \sim (\,\mathbf{0}, \mathbf{L}(t)\,) \tag{2.11}$$

The nearest neighbor structure of $\mathbf{L}(t)$ in (2.11) or (2.9) reflects a corresponding local structure to the statistical model for the estimated field error covariance. We will use this observation in Section 2.4 to develop tractable yet near optimal filtering algorithms.

2.3.2 Multi-Frame Case

Now we consider the multi-frame extension of the single-frame formulation given in (2.6),(2.7). The continuous optimization problem (2.3) can be considered to be an *optimal smoothing* problem based on the following temporal, linear Gauss-Markov dynamic system for $\mathbf{x}(t)$ [16]:

$$\frac{\partial}{\partial t}x(z_1, z_2, t) = q(t) \tag{2.12}$$

[1] More precisely, to avoid biasing the optical flow estimates towards zero, we only assume that the *gradients* of the optical flow field components are equal to the *gradients* of the Brownian motion processes. This avoids placing a constraint on the DC (i.e. average) value of the optical flow and focuses only on imposing a preference for smoothness in the flow.

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 (i.e. obtained by running a Kalman filter in each of the causal and anti-causal directions) are applicable [7]. In general we wish to compute only the most recent estimate $\hat{\mathbf{x}}(z_1, z_2, \tau)$ from (2.3) 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

$$\mathbf{x}(t) = \mathbf{x}(t-1) + \mathbf{q}(t), \quad \mathbf{q}(t) \sim \left(\mathbf{0}, \rho^{-1} I \right) \tag{2.13}$$

coupled with the observations given by (2.6),(2.7). The process noise $\mathbf{q}(t)$ is uncorrelated over time and captures the uncertainty in the dynamic model (2.13). This Gauss-Markov dynamic model, a discrete version of (2.12), indicates that the optical flow evolves in time as the accumulation of a random perturbation at each time frame. While we will be concerned with temporal dynamics of the form (2.13), naturally more complicated dynamic models, corresponding to different temporal coherence terms in (2.3), could be used.

2.4 Sequential Multi-Frame Estimation

In this section we consider state estimation for the dynamic system represented by (2.13),(2.6),(2.7). Conceptually, we may use well-developed optimal sequential estimation algorithms, such as the Kalman filter and its variants, for solution of this multi-frame optical flow estimation problem. One such algorithm, that will prove convenient for us, is the following implementation of the *information form* [13, 15] of the Kalman filter [16]:

- prediction stage

$$\bar{\mathbf{L}}(t) = \rho I - \rho^2 \left(\hat{\mathbf{L}}(t-1) + \rho I \right)^{-1} \tag{2.14}$$

$$\bar{\mathbf{x}}(t) = \hat{\mathbf{x}}(t-1) \tag{2.15}$$

$$\bar{\mathbf{z}}(t) = \bar{\mathbf{L}}(t)\bar{\mathbf{x}}(t) \tag{2.16}$$

- update stage

$$\hat{\mathbf{L}}(t) = \bar{\mathbf{L}}(t) + \mathbf{C}^T(t)\mathbf{W}(t)\mathbf{C}(t) + \mathbf{S}^T\mathbf{S} \tag{2.17}$$

$$\hat{\mathbf{z}}(t) = \bar{\mathbf{z}}(t) + \mathbf{C}^T(t)\mathbf{W}(t)\mathbf{y}(t) \tag{2.18}$$

$$\hat{\mathbf{L}}(t)\hat{\mathbf{x}}(t) = \hat{\mathbf{z}}(t) \tag{2.19}$$

where $\bar{\mathbf{x}}(t)$ is the one-step predicted estimate and $\hat{\mathbf{x}}(t)$ is the updated estimate using the new data available at time t. Also, $\bar{\mathbf{L}}(t)$ and $\hat{\mathbf{L}}(t)$ denote the predicted and updated information matrices, respectively. Note that the updated estimate $\hat{\mathbf{x}}(t)$ in (2.19) is specified implicitly, as for the single-frame case (2.9).

2.4.1 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 (2.14)–(2.19) impractical as the associated information matrices $\bar{\mathbf{L}}(t)$ and $\hat{\mathbf{L}}(t)$ of the optimal filter will 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. The suboptimal filtering algorithm presented below employs a systematic and rational approximation of Kalman filter, which is based on the propagation of approximate local *models* of the estimation error covariance, as discussed in connection with (2.11).

To develop our sub-optimal filter, consider the set of equations (2.14)-(2.19). First consider the update stage of the Kalman filter. If $\bar{\mathbf{L}}(t)$ possesses a sparse and banded nearest neighbor structure, as was true for the single-frame problem, then (2.17) will preserve this structure in $\hat{\mathbf{L}}(t)$ since, as we pointed out in connection with (2.9), $\mathbf{C}^T(t)\mathbf{W}(t)\mathbf{C}(t) + \mathbf{S}^T\mathbf{S}$ also possesses this structure. In particular, if this is the case, then (2.19) may still be solved efficiently for the updated estimate $\hat{\mathbf{x}}(t)$, and in fact this step would have *exactly* the same computational complexity as in the single-frame case. Thus, we desire to preserve such a sparse and banded structure in $\bar{\mathbf{L}}(t)$.

Now consider the prediction stage. Unfortunately, even if $\hat{\mathbf{L}}(t-1)$ in (2.14) is initially sparse and banded, the predicted information matrix $\bar{\mathbf{L}}(t)$ 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 in the present framework, then, is a sparse and banded *approximation* to $\bar{\mathbf{L}}(t)$ that may be efficiently computed.

As detailed in [16, 21], such an approximation may indeed be obtained by expanding the matrix inverse on the right hand of (2.14) in a series as follows:

$$\bar{\mathbf{L}}(t) = \rho I - \rho^2(\Omega^{-1} - \Omega^{-1}\Delta\Omega^{-1} + \Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1} - \cdots) \qquad (2.20)$$

where Ω is a block diagonal matrix whose 2×2 diagonal blocks are identical to the corresponding diagonal blocks of the matrix $\hat{\mathbf{L}}(t-1) + \rho I$ while $\Delta = \hat{\mathbf{L}}(t-1) + \rho I - \Omega$ is given by the remaining off-diagonal part of $\hat{\mathbf{L}}(t-1) + \rho I$. Note that Ω^{-1} is block diagonal. The series (2.20) 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 (2.14) by the following two-term approximation:

$$\bar{\mathbf{L}}(t) = \rho I - \rho^2(\Omega^{-1} - \Omega^{-1}\Delta\Omega^{-1}) \qquad (2.21)$$

Unlike (2.14), the suboptimal prediction step (2.21) does indeed preserve the desired nearest neighbor structure in the (approximated) information matrix $\bar{\mathbf{L}}(t)$.

It can be verified straightforwardly that propagating the information matrix in the approximate filter as in (2.17) and (2.21) 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 $O($ non-zero elements. Thus, the approximate filter has significant computation 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 $\overline{\mathbf{x}}(t)$), with the new observation. In particular, $\overline{\mathbf{L}}(t)$ can naturally be thought of as specifying a prior model for the error $\mathbf{e}(t) \equiv \mathbf{x}(t) - \overline{\mathbf{x}}(t)$ in the current estimate of the following form:

$$\overline{\mathbf{L}}(t)\,\mathbf{e}(t) = \zeta(t), \qquad \zeta(t) \sim \left(\mathbf{0}, \overline{\mathbf{L}}(t)\right) \tag{2.22}$$

which is just the counterpart of (2.11) 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 $\hat{\mathbf{x}}(t)$ in (2.19) is then equal to $\overline{\mathbf{x}}(t) + \hat{\mathbf{e}}(t)$. The update stage is thus just a static spatial estimation problem, where (2.22) represents a prior model just before the inclusion of new data. That is, by writing the observation equations (2.6),(2.7) concisely as $\mathbf{g}(t) = \mathbf{H}(t)\mathbf{x}(t) + \mathbf{r}(t)$, where $\mathbf{g}(t) \equiv [\mathbf{y}(t)^T, \mathbf{0}^T]^T$, $\mathbf{H}(t) \equiv [\mathbf{C}(t)^T, \mathbf{S}(t)^T]^T$, and $\mathbf{r} = [\mathbf{r}_1^T, \mathbf{r}_2^T]^T$, the estimate $\hat{\mathbf{e}}(t)$ can be obtained by solving the following static spatial estimation problem:

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{g}(t) - \mathbf{H}(t)\overline{\mathbf{x}}(t) \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{L}}(t) \\ \mathbf{H}(t) \end{bmatrix} \mathbf{e}(t) + \begin{bmatrix} -\zeta(t) \\ \mathbf{r}(t) \end{bmatrix}. \tag{2.23}$$

which is statistically equivalent to obtaining the updated estimate $\hat{\mathbf{x}}(t)$ of the unknown $\mathbf{x}(t)$ given the prediction $\overline{\mathbf{x}}(t)$ and observation $\mathbf{g}(t)$. Since the implicit model is specified by $\overline{\mathbf{L}}(t)$, our approximation of this matrix by a sparse matrix of the given nearest neighbor structure in (2.21) corresponds naturally to the specification of an approximate, *reduced-order* model for the spatial 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 [16]. Our approximation thus has a rational basis in estimation-theoretic considerations.

2.4.2 Numerical Experiments

We demonstrate the beneficial effects of the temporal coherence constraint, formulated as the dynamic model (2.13), and the efficacy of our near-optimal filter for optical flow estimation by numerical example in this section. Recall that, for images of realistic dimension, such as we consider here, exact implementation of the optimal Kalman filtering equations is impossible and thus we apply the temporal coherence constraint (2.13) via our suboptimal filter of Section 2.4.1.

A detailed comparison of the suboptimal and true optimal filters demonstrating the near-optimality of our approximation can be found in [16]. Here, a synthetic image sequence of a moving brightness pattern is processed by various multi-frame and single-frame optical flow estimation methods, and the improvements gained by using the particular temporal coherence constraint (2.13), as implemented by the filter we presented in Section 2.4.1, are compared to the conventional methods. Specifically, the following two methods are considered:

- **SF** (Single Frame)
  This method is a discrete version of the single-frame computational approach proposed by Horn and Schunck [9]. Each frame of optical flow is computed independently, i.e., without any provision for temporal integration of data, by solving the inversion problem (2.9) for $\hat{\mathbf{x}}(t)$.

- **TCS** (Temporal Coherence, Suboptimally computed)
  This method is the suboptimal but computationally efficient version of the optimal Kalman filter: as described in Section 2.4.1, the prediction step (2.14) of the Kalman filter is approximated as (2.21).

Variants of these methods arise in different computational environments. Specifically, the inversion steps (2.9) for **SF** and (2.19) for **TCS** can be implemented by one of the following computational procedures, leading to variations in the algorithms above:

- **ic** (iterative inversion, iterations to convergence)
  In practice, the inversion problems are solved iteratively. We use Gauss-Seidel iterations in the experiments here. Needless to say, this iterative solution will converge to the true solution 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 [9], although such estimates are suboptimal in the statistical sense.

In this section, each computational method is made explicit by the name of its main algorithm suffixed by the name of the variation, e.g., **TCS-ic**, **SF-is**, etc. Also, in each experiment, the initial frame of optical flow estimate is computed identically by the **SF-ic** method for every participating computational method in order to highlight the differences in the temporal effects of each method.

The method **SF-is** deserves special attention. This method is the approach to multi-frame optical flow estimation suggested by Horn and Schunck in [9]. It performs only one Gauss-Seidel iteration for the inverse problem (2.9) at each t but uses the estimate from the *previous* frame, $\hat{\mathbf{x}}(t-1)$, to initialize the current iteration. Unlike the **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.

One of the advantages of using a temporal coherence constraint in optical flow estimation is the improvement in the estimates due to the reduced effect of measurement noise through the averaging of the noisy data over time. Another, less obvious advantage, is the temporal accumulation of complementary information regarding the flow vectors. Reconstruction of optical flow using only spatial data integration (i.e., the **SF-is** method) 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. The example we give below demonstrates both the fact that the temporal constraint is instrumental in correctly estimating the flow in such cases and that it aids in noise suppression.

Stagnation Flow Experiment

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 2.2 shows a flow pattern whose velocity vector at point (z_1, z_2) is given by $(\alpha z_1, -\alpha z_2)$ for $\alpha = 0.1$, where the coordinate origin is at the midpoint of the bottom edge of the figure. This type of flow (for an arbitrary constant α) is useful for a local characterization of *stagnation flow* [25], 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 2.3 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.*

   A 9×9 unit uniform stencil is used to spatially smooth the images before brightness gradients are computed. The computational parameters $\rho = 400$ and $\nu = 40$ have been used. Figure 2.4 shows frame 18 of the estimated flow vectors computed by the **SF-ic** and **TCS-ic** methods. The

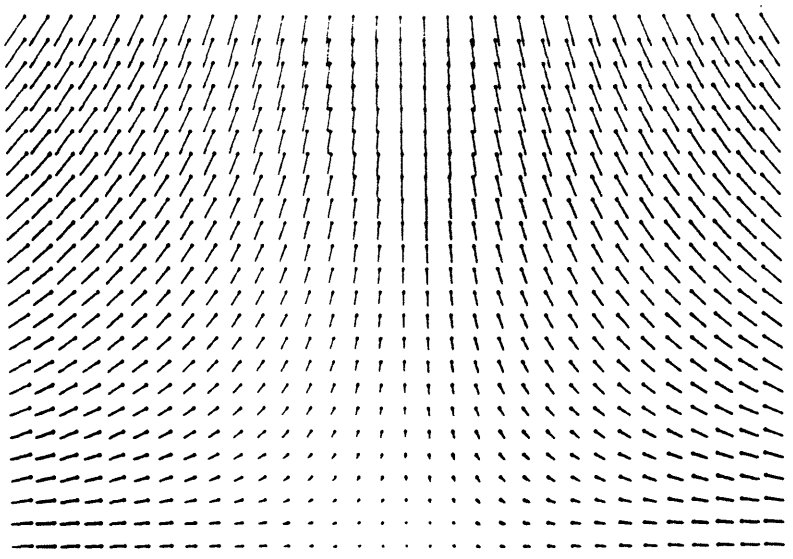

Figure 2.2: The true flow in the Stagnation Flow experiment. Every other flow vector along each axes is shown with a magnification factor of 4 for clarity.

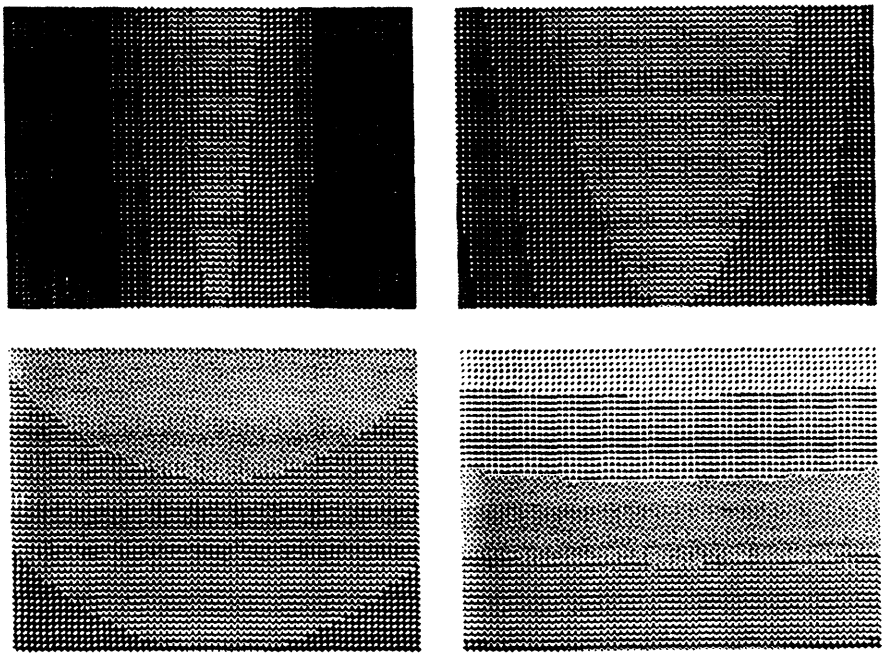

Figure 2.3: The Stagnation Flow image sequence. Frames 0 and 7 (top row) as well as 14 and 21 (bottom row) are shown.

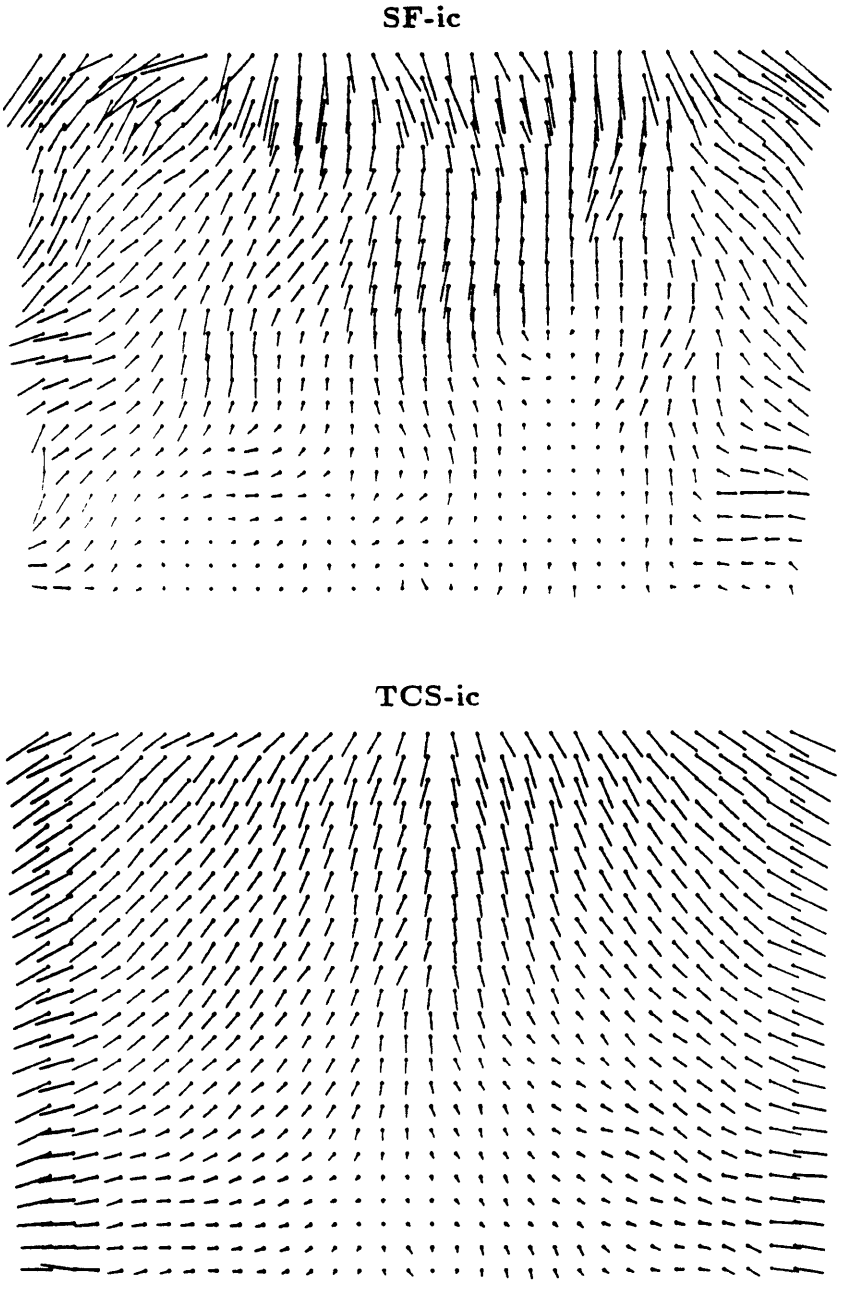

Figure 2.4: The optical flow estimates for frame 18 of the Stagnation Flow sequence by the **SF-ic** and **TCS-ic** methods.

SF-ic method, without any provision for temporal data integration, has completely failed to reconstruct the flow field, while the **TCS-ic** method has performed a reasonable reproduction of the flow in Figure 2.2. The flows computed by the **SF-is** and **TCS-is** are shown on Figure 2.5, which also displays the importance of temporal coherence in reconstruction. Figure 2.6 displays the percent average estimation error for each t.

$$\frac{\|\hat{\mathbf{x}}(t) - \mathbf{x}(t)\|^2}{\|\mathbf{x}(t)\|^2} \times 100. \tag{2.24}$$

where $\mathbf{x}(t)$ is the true flow and $\hat{\mathbf{x}}(t)$ is the estimated flow, for the four methods. These errors are consistent with our previous observations. Again, superior performance of the **TCS**-type methods over the **SF**-type methods is displayed rather dramatically by the error curves.

2.5 Multiscale Model-Based Estimation

One of the major computational bottlenecks of the Kalman filtering algorithm of Section 2.4 is the spatial estimation problem represented by (2.19). In Section 2.4.1 we were able to transform this step of the multi-frame problem back to the level of complexity of the single-frame Horn and Schunk formulation as a result of our reduced-order approximation to the field model (2.21), which resulted in a sparse and banded structure of the matrix $\hat{\mathbf{L}}(t)$. While such an approximation succeeds in making the multi-frame problem tractable while preserving near optimality of the resulting estimates and indeed represents, to our knowledge, the first implementation of the *complete* Kalman filtering equations to problems of this scale, the resulting inverse problem still leads to computationally intensive algorithms. Specifically, the associated set of equations (2.19), while sparse, is extremely large, corresponding to discrete versions of elliptic partial differential equations [9]. The standard approaches to solving such large sets of linear equations, such as the Gauss-Seidel [9, 26], multigrid [24], and successive over-relaxation (SOR) [27] algorithms, are iterative, requiring increasing numbers of iterations (and thus increasing per pixel computational load) as the image size grows.

In this section we examine a novel approach to such large, computationally intensive spatial estimation problems wherein we combine a *multiscale prior model* of the field with a set of field observations. Recall from Section 2.4.1 that we may view the update step of the optimal Kalman filter as such a static spatial estimation problem wherein a prior spatial field model (2.22) is combined with a set of observations $\mathbf{g}(t)$. The use of a multiscale modeling paradigm leads to extremely efficient estimation algorithms which hold the promise of *dramatically* reducing the computation required to solve such problems. For example, the resulting multiscale algorithm is not iterative and in fact requires a fixed number of floating point operations per pixel *independent of image size*. Thus, the computational savings associated with the new approach actually increases as the image size grows. For simplicity and to illustrate the issues involved we will focus only on the single-frame case here. For clarity we will

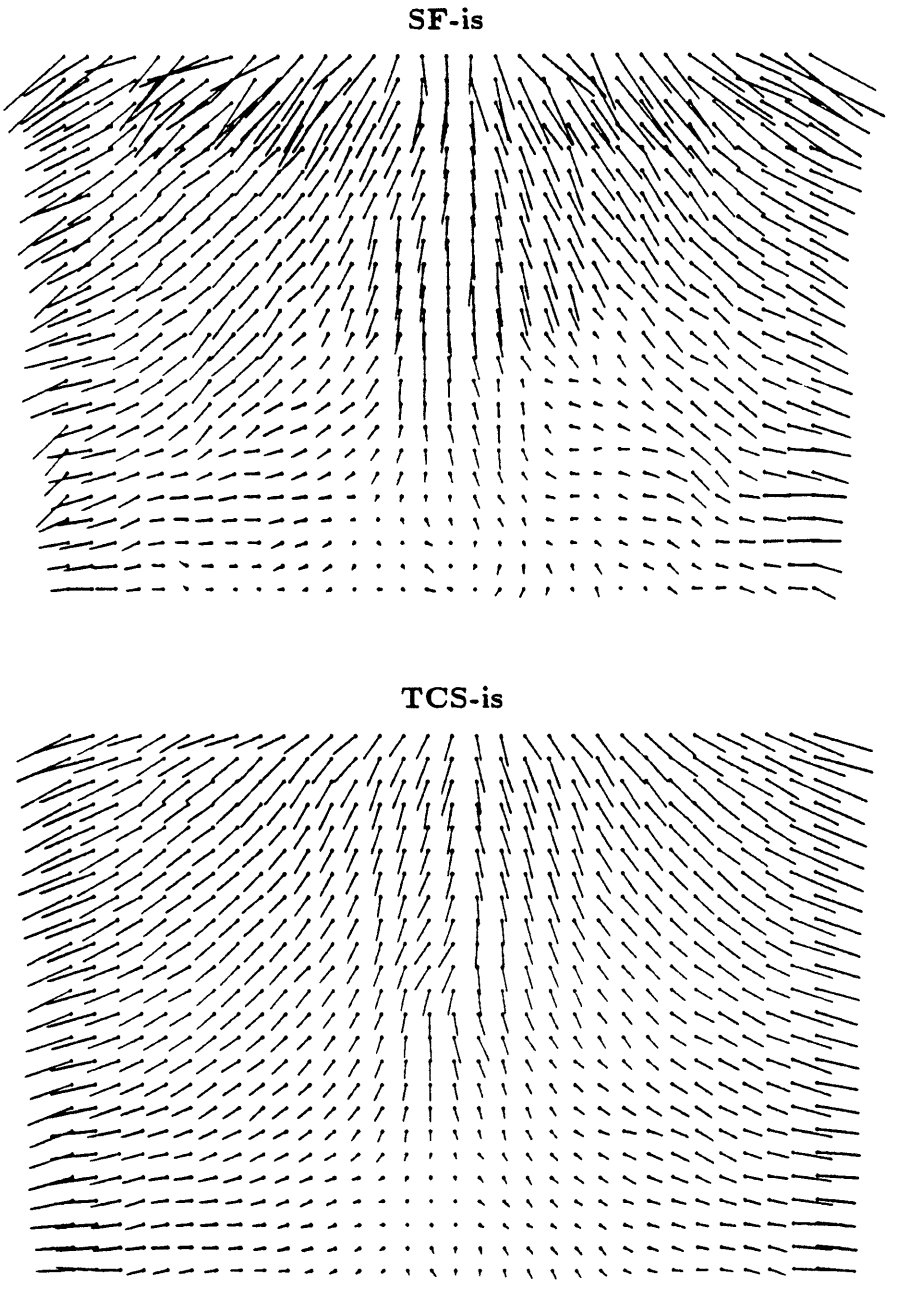

Figure 2.5: The optical flow estimates for the frame 18 of the Stagnation Flow sequence by the **SF-is** method and **TCS-is** method.

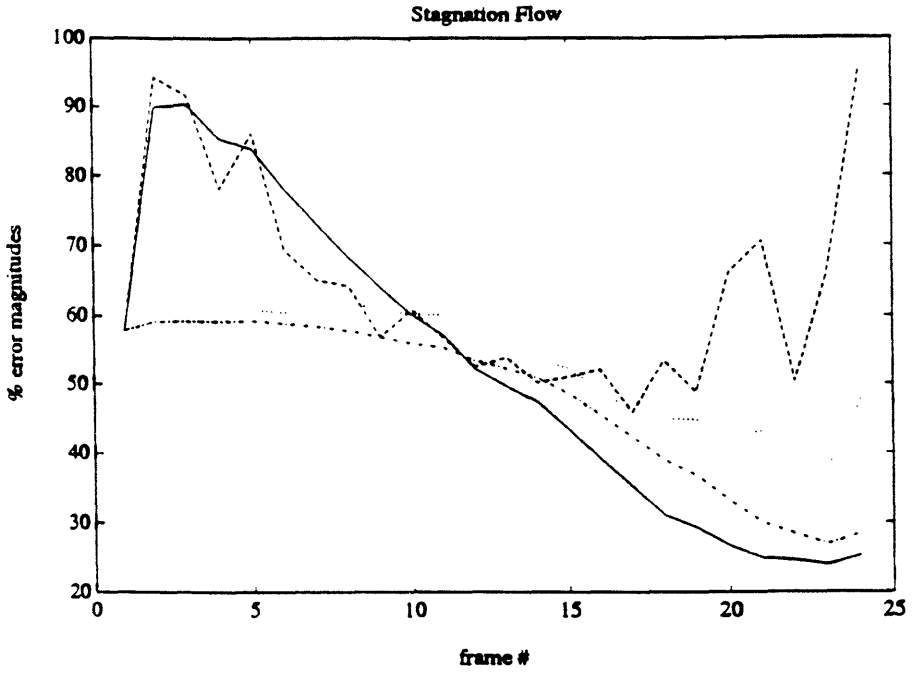

Figure 2.6: The estimation errors by the **TCS-ic** (solid-line), **SF-ic** (dashed-line), **SF-is** (dotted-line), and **TCS-is** (dash-dot line) methods for the Stagnation Flow experiment.

drop the time index t from the notation for the rest of this section, with the understanding that all quantities are taken at this point in time.

Recall, we argued in Section 2.3.1 that the single-frame optical flow estimates corresponding to (2.5) could be viewed as arising from the combination of the prior statistical spatial model (2.8), corresponding to the smoothness constraint, and the observations (2.6). Now, there is nothing special about the prior model (2.8) associated with the smoothness constraint. Thus we are lead to the idea of using a different class of prior models, capabable of capturing a wide range of phenomenon, and in particular of yielding behavior which is similar in nature to that corresponding to the smoothness constraint, but which lead to computationally more attractive problem formulations. That is, we want to change the smoothness constraint term $\mathbf{x}^T\mathbf{S}^T\mathbf{S}\mathbf{x}$ in (2.5) to something similar, say, $\mathbf{x}^T\mathbf{\Lambda}\mathbf{x} \approx \mathbf{x}^T\mathbf{S}^T\mathbf{S}\mathbf{x}$ (where $\mathbf{\Lambda}$ is a symmetric positive semi-definite matrix) such that the resulting optimization problem is easy to solve. If we factor $\mathbf{\Lambda}$ as $\mathbf{\Lambda} = \mathbf{S}^T\bar{\mathbf{S}}$ then we can interpret the new constraint as a prior probabilistic model just as we did with the smoothness constraint. In addition, there is a precise interpretation of what we have done as a Bayesian estimation problem. Specifically, if $\mathbf{\Lambda}$ is invertible, then the use of this new constraint in place of the smoothness constraint is equivalent to modeling the flow field probabilistically as $\mathbf{x} \sim (\mathbf{0}, \mathbf{\Lambda}^{-1})$, since in this case the Bayes' least squares estimate of the flow

field $\mathbf{x}$, given this prior model and the measurements in (2.6) is provided by:

$$\hat{\mathbf{x}}_{BLSE} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ (\mathbf{y} - \mathbf{C}\mathbf{x})^T \mathbf{W} (\mathbf{y} - \mathbf{C}\mathbf{x}) + \mathbf{x}^T \mathbf{\Lambda} \mathbf{x} \right\} \tag{2.25}$$

The normal equations corresponding to (2.25) are given by:

$$(\mathbf{C}^T \mathbf{W} \mathbf{C} + \mathbf{\Lambda}) \hat{\mathbf{x}}_{BLSE} = \mathbf{C}^T \mathbf{W} \mathbf{y} \tag{2.26}$$

Comparison of the problem formulations (2.5) and (2.25), or of the normal equations (2.9) and (2.26), makes it apparent how the two problem formulations are related. The choice of the new prior model corresponding to $\mathbf{\Lambda}$ is now clearly at the heart of the problem. We introduce our class of new models next.

2.5.1 A Class of Multiscale Models

The models we utilize to replace the smoothness constraint prior model were recently introduced in [28, 29, 30, 31]. These models represent the flow field at multiple scales, i.e. for a set of scales $m = 0, \ldots, M$, with $m = 0$ being the coarsest scale and $m = M$ the finest scale, we define a set of optical flow fields indexed by scale and space, namely $x_m(i,j)$. At the m-th scale, the field consists of 4^m flow vectors, as illustrated in Figure 2.7, capturing features of the optical flow field discernible at that scale (i.e. finer-resolution features of the field appear only in finer-scale representations). Thus, the coarsest version of the flow field consists of just a single vector, corresponding to the average value of the optical flow over the entire spatial domain of interest, and successively finer versions consist of a geometrically increasing number of vectors. At the finest level, the flow field is represented on a grid with the same resolution as the image brightness data. In particular, $x_M(i,j)$ corresponds to the optical flow vector $x(i,j)$ in (2.4).

Abstractly, we are representing the flow field on the *quadtree structure* illustrated in Figure 2.8. Pyramidal data structures such as the quadtree naturally arise in image processing algorithms which have a multiscale component. For instance, successive filtering and decimation operations lead to images defined on such a hierarchy of grids in the Laplacian pyramid coding algorithm of Burt and Adelson [32] and in the closely related wavelet transform decomposition of images [33]. Also, the multigrid approaches to low level vision problems discussed by Terzopoulos [24] involve relaxation on a similar sequence of grids.

The model we introduce in this section describes in a probabilistic manner how the optical flow field $x(i,j) = x_M(i,j)$ is constructed by adding detail from one scale to the next. Just as the smoothness constraint prior model (2.8) describes probabilistic constraints among values of the optical flow at different spatial locations, our multiscale model describes such constraints among values at different *scales*. That is, our model describes the probabilistic evolution of $x_m(i,j)$ as the scale m evolves from coarse to fine. For notational convenience in describing such models, we denote nodes on the quadtree with a single abstract index s which is associated with the 3-tuple (m,i,j) where, again, m is the scale and (i,j) is a spatial location in the grid at the m-th scale. It is also useful to

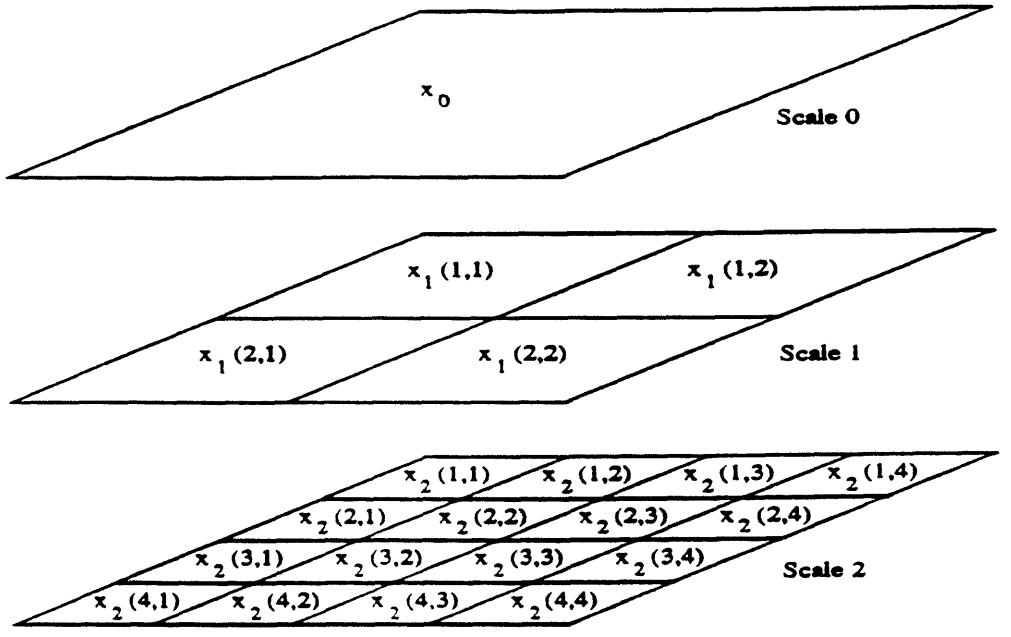

Figure 2.7: The structure of a multiscale optical flow field is depicted. The components of the field are denoted $x_m(i,j)$ where m refers to the scale and the pair (i,j) denotes a particular grid location at a given scale. At the coarsest scale, there is a single flow vector and, more generally, at the m-th scale there are 4^m vectors.

define an *upward shift operator* $\bar{\gamma}$. In particular, the *parent* of node s is denoted $s\bar{\gamma}$ (see Figure 2.8). We note that the operator $\bar{\gamma}$ is not one-to-one: it is in fact four-to-one since each node will have four "offspring" at the next scale. For instance, if s corresponds to any of the nodes in the upper left quadrant of the second level grid (see Figure 2.7), i.e. nodes $(2,1,1),(2,2,1),(2,1,2)$ or $(2,2,2)$, then $s\bar{\gamma}$ corresponds to their parent on the first level, namely node $(1,1,1)$.

We are now in a position to define the class of multiscale models which describe the evolution of a multiscale stochastic processes indexed by nodes on the quadtree. Specifically, a stochastic quadtree process $x(s)$ is described recursively by:

$$x(s) = A(s)x(s\bar{\gamma}) + B(s)w(s) \tag{2.27}$$

under the following assumptions:

$$x_0 \sim (\mathbf{0}, P_0) \tag{2.28}$$

$$w(s) \sim (\mathbf{0}, I) \tag{2.29}$$

The vectors $x(s)$ and $w(s)$ are referred to as the state and driving noise terms. The state variable x_0 at the root node of the tree provides an initial condition for the recursion. The driving noise is white in both space and scale, and is uncorrelated with the initial condition. Interpreting each level as a representation of a two-dimensional field, we see that (2.27) describes the evolution of the

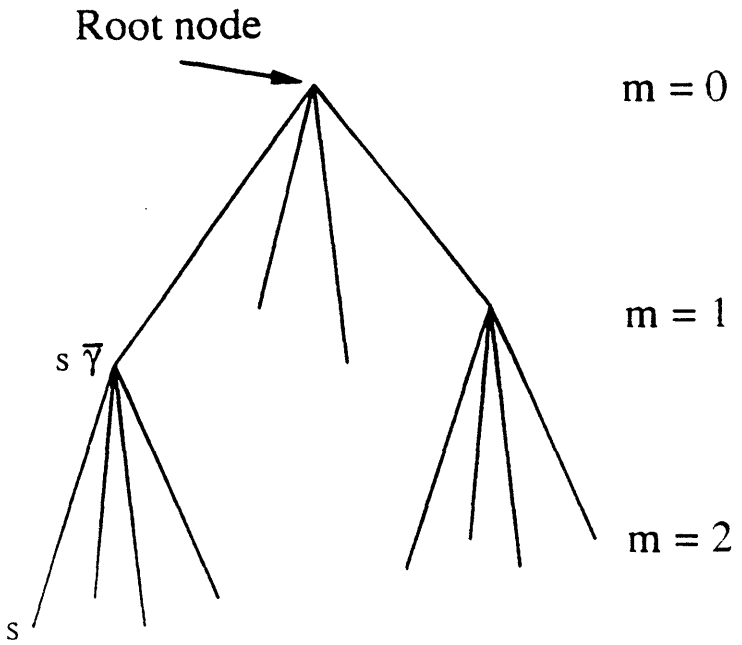

Figure 2.8: Quadtree structure on which the multiscale processes are defined. The abstract index s refers to a node in the quadtree: $s\bar{\gamma}$ refers to the parent of node s.

process from coarse to fine scales. The term $A(s)x(s\bar{\gamma})$ represents interpolation down to the next level, and $B(s)w(s)$ represents higher resolution detail added as the process evolves from one scale to the next. In the application of interest here, $x(s) = x_m(i,j)$, where $s = (m,i,j)$, and thus $A, B \in \Re^{2\times 2}$. Such a model corresponds in essence to a first-order recursion in scale for optical flow.

Measurements of the finest level optical flow field are available from the brightness constraint. In particular, at the grid point (i,j) at the finest level M, we have a measurement equation corresponding to that in (2.4):

$$y(i,j) = C(i,j)\,x_M(i,j) + v(i,j) \qquad v(i,j) \sim \left(0, \nu^{-1}(i,j)\right) \tag{2.30}$$

where $C(i,j) \in \Re^{1\times 2}$ is the spatial brightness gradient at location (i,j) and the white Gaussian observation noise is assumed to be independent of the initial condition x_0 and the driving noise $w(s)$ in (2.27)–(2.29). Of course, we can group the state variables $x(s)$ at the finest level into a vector $\mathbf{x}_M$ as well as the corresponding measurements $y(s)$ and spatial gradient terms $C(s)$ in the same way as we did to get (2.6):

$$\mathbf{y} = \mathbf{C}\,\mathbf{x}_M + \mathbf{v} \qquad \mathbf{v} \sim \left(\mathbf{0}, \mathbf{W}^{-1}\right) \tag{2.31}$$

We now have exactly the framework which led to the statement of (2.25) as an alternative to the smoothness constraint formulation (2.5). In particular, the modeling equations (2.27)–(2.29) indicate that at the finest level of the

quadtree, the flow field vectors will be a set of jointly Gaussian random variables $\mathbf{x}_M \sim (\mathbf{0}, \Lambda^{-1})$, where Λ^{-1} is implicitly given by the parameters in (2.27)–(2.29). The Bayes' least squares estimate of $\mathbf{x}_M$ given the measurements in (2.31) and the prior model (2.27)–(2.29) is then given by:

$$\hat{\mathbf{x}}_M = \underset{\mathbf{x}_M}{\operatorname{argmin}} (\mathbf{y} - \mathbf{C}\mathbf{x}_M)^T \mathbf{W} (\mathbf{y} - \mathbf{C}\mathbf{x}_M) + \mathbf{x}_M^T \Lambda \mathbf{x}_M \qquad (2.32)$$

The multiscale modeling framework thus provides an alternative to the smoothness constraint formulation of (2.5).

What remains to be done is to specify a model within this class that has characteristics similar to those of the smoothness constraint prior model. In particular, for our multiscale model based on (2.27)-(2.29) to approximate the smoothness constraint prior we would like to choose our model parameters so that we have $\mathbf{S}^T\mathbf{S} \approx \Lambda$. The observation in Section 2.3.1 that the prior model (2.8) implied by the operator S in (2.5) corresponds to a Brownian motion "fractal prior" suggests one approach to choosing the model parameters. In particular, the one-dimensional Brownian motion has a $1/f^2$ generalized spectrum [22]. It has been demonstrated that such processes are well approximated by multiscale models such as ours in one dimension if geometrically decreasing powers of noise are added at each level m of the process [30, 34]. In particular, this motivates the choice of $B(s) = b4^{-\mu m(s)} I$ in (2.27), where b and μ are scalar constants. The constant b directly controls the overall noise power in the process. Also, as discussed in [34], the choice of μ controls the power law dependence of the generalized spectrum of the process at the finest resolution as well as the fractal dimension of its sample paths. Specifically, this spectrum has a $1/f^{2\mu}$ dependence. Thus, the choice of $\mu = 1$ would correspond to a Brownian-like fractal process. To achieve greater flexibility in both the modeling and estimation, we allow μ to be a parameter that can be varied. In addition, recall that in the smoothness constraint formulation, $\mathbf{S}^T\mathbf{S}$ was not invertible because of the implicit assumption of infinite prior variance on the DC value of the optical flow field. In our multiscale regularization context, this would correspond to setting the initial covariance P_0 equal to infinity in (2.28). This can be done without difficulty in the estimation algorithms described next, but we have found that it is generally sufficient to simply choose P_0 to be a large multiple of the identity.

We have now specified a class of models which will allow us to approximate the smoothness constraint prior model. The simple multiscale structure of these models leads to very efficient algorithms for computing the optimal estimate of the state given a set of measurements. One of these algorithms, which we refer to as the Multiscale Regularization (MR) algorithm, was developed in [28, 29, 30, 35] for one-dimensional signals, and its extension to images is described in [36].

The MR algorithm computes the Bayes least squares estimate of the state vectors (2.27) given the measurements (2.30) in two steps. The first step is an *upward* or *fine-to-coarse* sweep on the quadtree, which propagates the measurement information in parallel, level by level, from the fine scale nodes up to the root node. This step produces the best estimate at each node given all the data in the subtree under that node. At the top of the tree, one obtains the

smoothed estimate of the root node, that is, the estimate of this node based on *all* of the data. The smoothed estimate and associated error covariance at the root node then provide initialization for the next step. This second step is a *downward* or *coarse-to-fine* sweep which propagates the global measurement information now at the root node back down, and throughout the tree. The result at each node is the least squares estimate $\hat{x}^s(s)$ of the state $x(s)$ based on all of the data. The resulting estimates at the finest level of the quadtree provide the solution to (2.32). The resulting algorithm is just a generalization of the Rauch-Tung-Striebel smoothing algorithm [37] *in scale*. The details of the upward and downward sweeps are discussed in greater detail in [30, 35, 36].

2.5.2 Numerical Experiments

Here we demonstrate the substantial computational benfit that can be achieved through the use of our multiscale modeling pardigm. To specify the MR algorithm completely we need to choose the parameters of the model. We utilize the following parameterization:

$$x(s) = x(s\bar{\gamma}) + (b4^{-\mu m(s)})w(s) \quad (2.33)$$

$$y(s) = C'(s)x(s) + v(s) \quad (2.34)$$

$$w(s) \sim (\mathbf{0}, I) \quad (2.35)$$

$$v(s) \sim (0, \nu^{-1}(s)) \quad (2.36)$$

$$x_0 \sim (\mathbf{0}, pI) \quad (2.37)$$

From (2.33) and (2.35) we see that the two components of the optical flow field are modeled as independent sets of random variables, and that each will have a fractal-like characteristic due to the choice of the driving noise gain $B(s)$ (as discussed in the previous section). We view μ and b as free model parameters which can be varied to control the degree and type of regularization in much the same way that the parameter ν in the smoothness constraint formulation (2.2) is used to tradeoff between the data dependent and regularization terms in the optimization functional.

As discussed previously, the measurements $y(s)$ and measurement matrix $C'(s)$ come directly from the image temporal and spatial gradients, which are available at the finest level of the quadtree. In the experiments described below, we use a simple two-image difference to approximate the temporal gradient. The spatial gradient is computed by smoothing the image with a 3×3 Gaussian kernel followed by a central difference approximation. The additive noise variance is given by $\nu^{-1}(s)$. We have found empirically that the choice $\nu^{-1}(s) = max(\|C'(s)\|^2, 10)$ works well. This choice effectively penalizes large spatial gradients, which are likely points of occlusion where the brightness constraint equation will not hold [38]. The parameter p in the prior covariance of the root node was set to $p = 100$. The distribution (2.37) on the root node effectively says that we are modeling the optical flow field components as zero mean random processes. The prior covariance reflects our confidence in this assumption. Since we do not believe that *any* prior assumption on the mean of optical flow field components can be justified, we set the parameter p such that

Figure 2.9: First frame of "Yosemite" sequence.

the implied standard deviation is much larger than the sizes of the flow fields we expect to see.

We compare our approach computationally and visually to the the Gauss-Seidel (GS) and successive over-relaxation (SOR) algorithms, which can be used to compute the solution of the smoothness constraint formulation given by (2.5). Straightforward analysis shows that the GS and SOR algorithms require 14 and 18 floating point operations (flops) per pixel per iteration respectively. The number of iterations required for convergence of the iterative algorithms grows with image size [27]. For reasonable size images (say, 512×512), SOR may require on the order of hundreds of iterations to converge, so that the total computation per pixel can be on the order of $10^3 - 10^4$ flops. On the other hand, the MR algorithm requires 76 flops per pixel. Note that the MR algorithm is *not* iterative. Thus, the computational gain associated with the MR algorithm can be on the order of one to two orders of magnitude. Details may be found in [36].

Yosemite Sequence Experiment

This example is a synthetic 256×256 image sequence which simulates the view obtained by flying through the Yosemite Valley[2]. The first image in the sequence is shown in Figure 2.9 along with the actual flow field in Figure 2.10. The flow computed via the MR algorithm is shown in Figure 2.11 and the smoothness constraint solution is shown in Figure 2.12. The smoothness constraint flow

[2]This sequence was synthesized by Lyn Quam of SRI International.

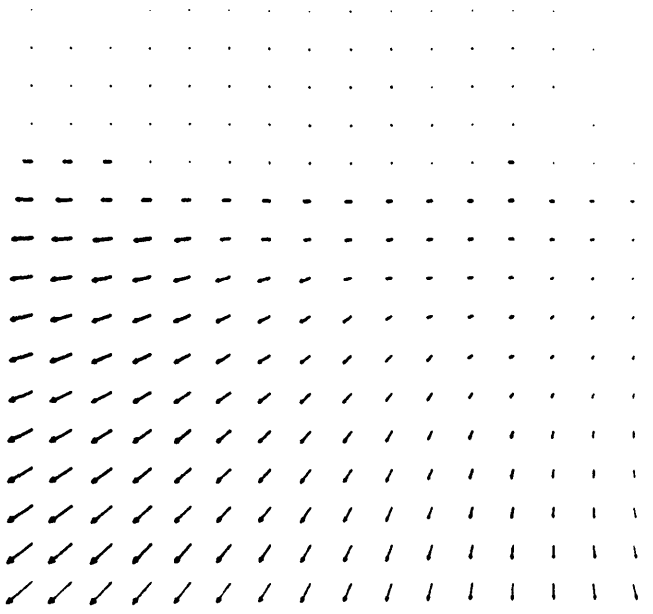

Figure 2.10: Yosemite sequence true optical flow.

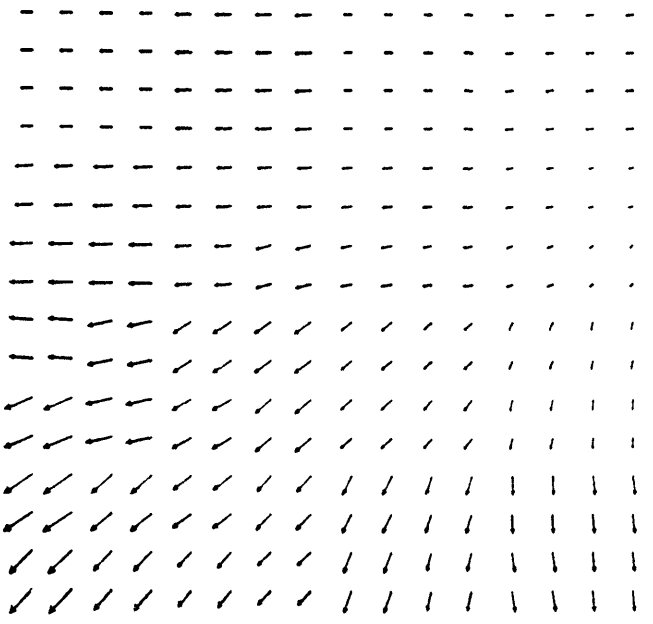

Figure 2.11: MR algorithm flow estimates: $b = 10, \mu = 2.5$.

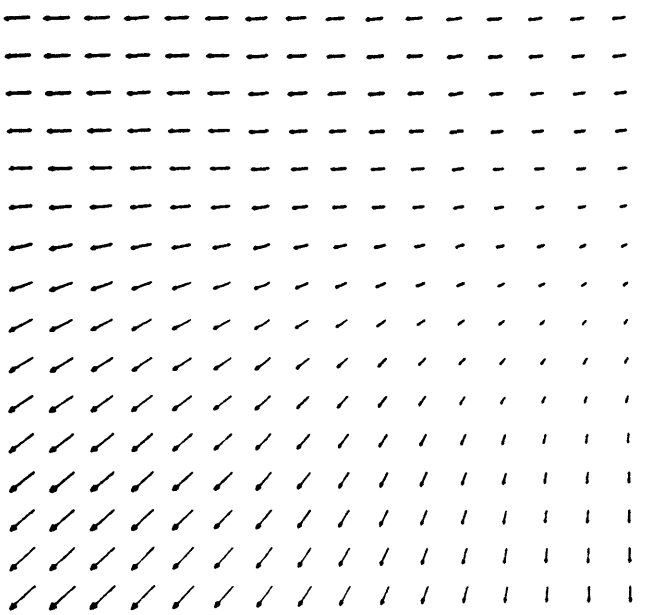

Figure 2.12: SOR algorithm flow estimates: $R = 500^2$, 250 iterations. The SOR algorithm required a factor of 60 more computation in this example.

estimates required 250 SOR iterations in this example, representing a factor of 60 more computation than the MR estimates. Note the substantial increase over the previous example in the number of iterations required for the SOR algorithm to converge. The number of iterations required for convergence depends on several things, including the parameter ν, the image gradient characteristics and the image size. Theoretical analysis in [27] shows that the SOR algorithm requires on the order of n iterations for an $n \times n$ image. Thus, we expect substantially more computational savings as the image size increases.

A root mean square (rms) error comparison of the algorithms is shown in Figure 2.13. As expected, the SOR algorithm is significantly faster than the GS algorithm (they will converge to the same result since they are solving the same partial differential equation). The rms error in the MR flow estimates is depicted as a straight line, since the algorithm is not iterative. Neither of the estimates coincides with the actual optical flow, but they do have comparable rms error as in the previous example. In addition, the figure illustrates the computational advantage of the MR algorithm. In particular, the SOR algorithm is still reducing the rms error in its flow estimates after 300 iterations, at which point the MR algorithm requires a factor of $300/4.2 = 71.4$ less computation.

This image sequence contains a problem often encountered in real images: regions of constant intensity. The problem is that the lack of gradient information in that region implies that the optical flow is not well defined. The smoothness constraint and multiscale prior models provide a means of interpolating out into these regions. The result of this is apparent in the top portion of Figures 2.11

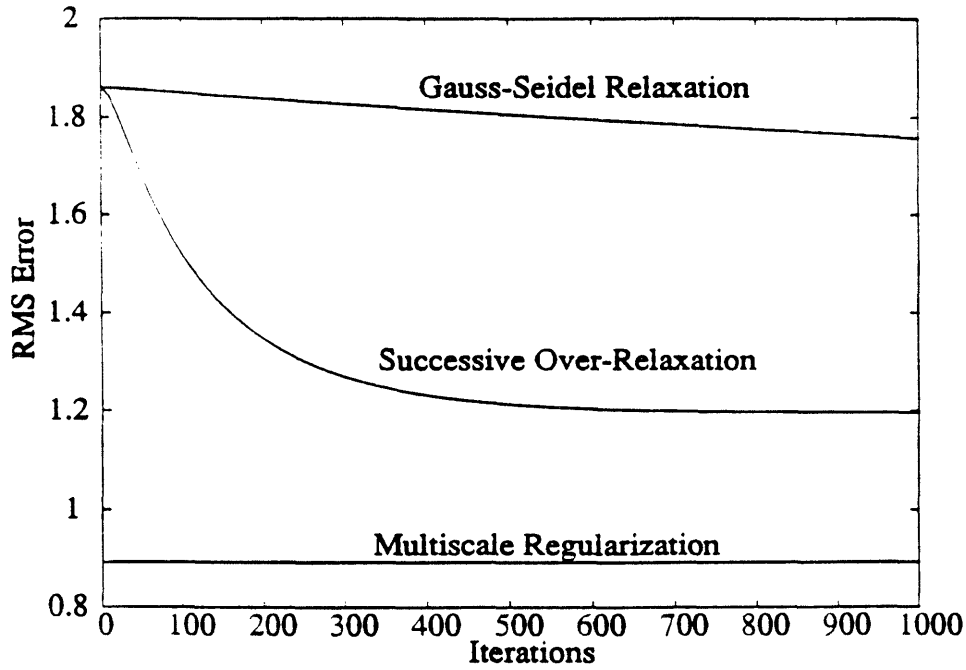

Figure 2.13: Rms Error Comparison of MR, SOR and Gauss-Seidel (GS) algorithm flow estimates for the Yosemite sequence.

and 2.12. An advantage of the MR formulation is that it accomplishes this extrapolation at an appropriately coarser, and hence computationally simpler, scale.

Note the the MR and SC flow estimates are not identical due to differences in the prior models. If there is particular interest in obtaining the SC solution, the question arises of using the MR solution as an initial guess for the iterative algorithms which compute the SC solution. Note that the difference between the SC and MR flow estimates is associated with the non-smooth, high frequency aspects of the MR flow at block edges. It is precisely these high frequency components that are quickly removed by SOR or GS algorithms computing the the smoothness constraint solution and suggests that the MR algorithm would provide an excellent *pre-conditioner* for the iterative algorithms.

2.6 Conclusions

We have taken an estimation-theoretic perspective to image sequence processing using as our vehicle the computation of optical flow. Our results made use of the interpretation of various components of standard formulations of this problem as *statistical models*. First we presented a near-optimal Kalman filter for the estimation of optical flow under a temporal coherence constraint and based on the propagation of approximate local *models* of the estimation error covariance. This filter provides, to our knowledge, the first implementation of the *complete* Kalman filtering equations for space-time problems of this scale, and the only example of successful, near optimal, propagation of covariance matrices of this size.

Next we used the observation that both the single and multi-frame problems can be formulated as spatial estimation problems, wherein sets of observations are fused with prior spatial field models, to motivate the use of a recently developed class of *multiscale* statistical models in their solution. The algorithms arising from this class of models allows *dramatic* speedups in computational speed. For simplicity we concentrated on the single frame case, though we could also use for the static spatial estimation problem occuring in the multi-frame situation. A particularly interesting and open question is how to directly propagate one of the multiscale models used in Section 2.5 in time.

Acknowledgements

This work was partially supported by the Charles Stark Draper Laboratory IR&D Program under Grant DL-H-418524, the Office of Naval Research under Grant N00014-91-J-1004, the National Science Foundation under Grant MIP-9015281, and the Army Research Office under Grant DAAL03-86-K-0171.

Bibliography

[1] 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*, pp. 238–245, IEEE Computer Society Press, 1990, Osaka, Japan.

[2] 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.

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

[4] S. Negahdaripour and B. K. Horn, "Direct passive navigation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, 1987.

[5] N. Baaziz and C. Labit, "Multigrid motion estimation on pyramidal representations for image sequence coding," Tech. Rep. 572, IRISA, Feb. 1991.

[6] D. Walker and K. Rao, "Improved pel-recursive motion compensation," *IEEE Transactions on Communications*, vol. 32, pp. 1128–1134, 1984.

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

[8] A. Rougee, B. Levy, and A. S. Willsky, "Reconstruction of two dimensional velocity fields as a linear estimation problem," in *Proceedings 1st International Conference on Computer Vision*, (London, England), pp. 646–650, 1987.

[9] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

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

[11] A. Singh. "Incremental estimation of image-flow using a Kalman filter." in *Proceedings of Workshop on Visual Motion*. pp. 36–43. IEEE Computer Society Press. 1991. Princeton. NJ.

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

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

[14] A. Gelb. ed.. *Applied Optimal Estimation*. Cambridge. MA: MIT Press. 1974.

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

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

[17] M. Bertero. T. Poggio. and V. Torre. "Ill-posed problems in early vision." *Proceedings of the IEEE*. vol. 76. pp. 869–889. 1988.

[18] E. C. Hildreth. "Computations underlying the measurement of visual motion." *Artificial Intelligence*. vol. 23. pp. 309–354. 1984.

[19] 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*. pp. 148–155. IEEE Computer Society Press. 1989. Irvine. CA.

[20] L. H. Matthies. R. Szeliski. and T. Kanade. "Kalman filter-based algorithms for estimating depth from image sequences." *International Journal of Computer Vision*. vol. 3. 1989.

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

[22] B. Mandelbrot and H. V. Ness. "Fractional Brownian motions. fractional noises and applications." *SIAM Review*. vol. 10. pp. 422–436. 1968.

[23] B. C. Levy. M. B. Adams. and A. S. Willsky. "Solution and linear estimation of 2-D nearest-neighbor models." *Proceedings of the IEEE*. vol. 78. pp. 627–641. 1990.

[24] D. Terzopoulos. "Image analysis using multigrid relaxation models." *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. PAMI-8. pp. 129–139. 1986.

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

[26] G. Strang. *Introduction to Applied Mathematics*. Wellesley. MA: Wellesley-Cambridge Press. 1986.

[27] C.-C. J. Kuo, B. C. Levy, and B. R. Musicus, "A local relaxation method for solving elliptic PDE's on mesh connected arrays," *SIAM J. Sci. Stat. Comput.*, vol. 8, pp. 550–573, 1987.

[28] K. C. Chou, A. S. Willsky, A. Benveniste, and M. Basseville, "Recursive and iterative estimation algorithms for multiresolution stochastic processes," in *Proc. of the IEEE Conference on Decision and Control*, Dec. 1989.

[29] K. C. Chou, *A stochastic modeling approach to multiscale signal processing*, PhD thesis, Massachusetts Institute of Technology, 1991.

[30] K. C. Chou, A. S. Willsky, and A. Benveniste, "Multiscale recursive estimation, data fusion and regularization," submitted to IEEE Transactions on Automatic Control, 1992.

[31] S. C. Clippingdale and R. G. Wilson, "Least squares image estimation on a multiresolution pyramid," in *Proc. of the 1989 International Conference on Acoustics, Speech, and Signal Processing*, 1989.

[32] P. Burt and E. Adelson, "The Laplacian Pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, pp. 482–540, 1983.

[33] S. Mallat, "Multi-frequency channel decomposition of images and wavelet models," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 2091–2110, 1989.

[34] G. Wornell, "A Karhunen-Loeve like expansion for 1/f processes," *IEEE Transactions on Information Theory*, vol. 36, pp. 859–861, 1990.

[35] K. C. Chou, A. S. Willsky, and R. Nikoukhah, "Multiscale systems, Kalman filters and Riccati equations," submitted to IEEE Transactions on Automatic Control, 1992.

[36] M. R. Luettgen, W. C. Karl, and A. S. Willsky, "Optical flow computation via multiscale regularization," submitted to IEEE Transactions on Image Processing, 1992.

[37] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, vol. 3, pp. 1445–1450, 1965.

[38] E. Simoncelli, E. Adelson, and D. Heeger, "Probability distributions of optical flow," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Maui, Hawaii), June 1991.